



11-01-11



Attorney Docket No.: 119941-1083

PATENT



## IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

"Express Mail" mailing label no. EL152182301US

In re Application of: Jeff L. DeJong

Date of Deposit October 30, 2000

Divisional of Serial No.: 09/326,529

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10, with sufficient postage, on the date indicated above and is addressed to Commissioner for Patents, Washington, D.C. 20231.

Filed: June 4, 1999

Linda L. Gibson  
(Printed Name)

Group Art Unit: 1652

Examiner: Richard Hutson

For: TRANSCRIPTION FACTORS RELATED TO TFIIA

BOX: NEW APPLICATION  
Commissioner for Patent  
Washington, D.C. 20231REQUEST FOR FILING DIVISIONAL APPLICATION  
UNDER 37 C.F.R. 1.53(b)

Dear Sir:

This is a request for filing a divisional application under 37 C.F.R. 1.53(b), of pending prior application 09/326,529, filed June 4, 1999, entitled TRANSCRIPTION FACTORS RELATED TO TFIIA. The inventor of the invention being claimed in this Application is: Jeff L. DeJong. Please amend the specification by inserting before the first line the sentence: "This application is a division of pending application number 09/326,529, filed on June 4, 1999."

1. A preliminary amendment is submitted herewith.
2. A true copy of the prior application as filed in the parent application is enclosed.
3. Informal drawings are enclosed.
4. A Declaration Claiming Small Entity Status under 37 C.F.R. 1.9 was filed in the prior application and such status is still proper and desired and a copy is enclosed herewith.

5. A Declaration is enclosed.
6. An Election Under 37 C.F.R. §§ 3.71 and 3.73 and Power of Attorney is enclosed.
7. A copy of the Sequence Listing with accompanying floppy disk is enclosed.
8. Please cancel in this application original Claims 1 and 3-32 of the prior application and add new claims 71-85 before calculating the filing fee.

The filing fee is calculated below:

For	Number Filed	Number Extra	Rate	Fee
Total Claims	54	34	9.00	\$306.00
Independent Claims	10	7	40.00	\$280.00
Basic Filing Fee				\$355.00
Multiple Dependent Claim	0	0	130.00	\$0.00
<b>TOTAL FILING FEE</b>				<b>\$941.00</b>

9. Two checks one in the amount of \$355.00 for the filing fee and one in the amount of \$586.00 for the extra claims are enclosed for the filing fee.

I hereby verify that the attached papers are a true copy of prior application Serial No. 09/326,529, filed June 4, 1999, and further that all statements made herein are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Please address all future correspondence to:

Gardere & Wynne, L.L.P.  
1601 Elm Street, Suite 3000  
Dallas, Texas 75201

and direct all telephone calls to Edwin S. Flores at (214) 999-4559.

Respectfully submitted,

GARDERE & WYNNE L.L.P.

By:   
Edwin S. Flores  
Attorney for Applicant  
Registration No.: 38,453

Date: October 30, 2000  
3000 Thanksgiving Tower  
1601 Elm Street  
Dallas, Texas 75201  
(214) 999-4559

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:

Jeff DeJong

§

§

Examiner:

Serial No.:

§

§

Group Art Unit:

Filed: Concurrently herewith

§

§

Atty. Dkt.: 119941-1053

For: Transcription Factors Related to TFIIA

§

**DECLARATION CLAIMING SMALL ENTITY STATUS  
37 C.F.R. §§ 1.9(f) and 1.27(d) - NONPROFIT ORGANIZATION**

Commissioner of Patents  
and Trademarks  
Washington, D.C. 20231

Sir:

I hereby declare that I am an official empowered to act on behalf of the nonprofit organization identified below:

Name of Organization: Board of Regents  
The University of Texas System

Address of Organization: 201 West 7th Street  
Austin, Texas 78701

The type of organization is a university.

I hereby declare that the organization identified above qualifies as a nonprofit organization as defined in 37 C.F.R. § 1.9(e) (1), and thus is a "small entity" as defined in § 1.9(f), for purposes of paying reduced fees under Sections 41(a) and (b) of Title 35, United States Code, with regard to the above-referenced application.

I hereby declare that exclusive rights to the invention have been conveyed to and remain with the organization, with respect to the above-referenced invention, nor have I assigned, granted, conveyed or licensed and am under no obligation under contract or law to assign, grant, convey or license, any rights in the invention to any person who could not be classified as an independent inventor under 37 CFR § 1.9(c) if that person had made the invention, or to any concern which would not qualify as a small business concern

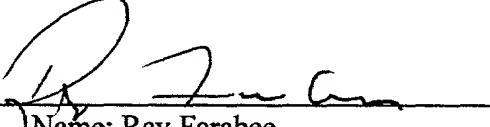
under 37 CFR § 1.9(d) or a nonprofit organization under 37 CFR § 1.9(e), with the exception that the Government may have rights in the invention pursuant to a funding agreement under 35 U.S.C. § 202(c)(4):

NONE

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application, any patent issuing thereon, or any patent to which this verified statement is directed.

BOARD OF REGENTS,  
THE UNIVERSITY OF TEXAS SYSTEM

By: 

Name: Ray Farabee  
Title: Vice Chancellor  
and General Counsel

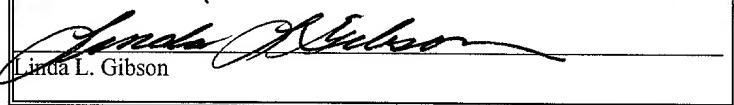
Date: May 27, 1999

719942.1

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re Application of: Jeff L. DeJong  
Divisional Serial No.: 09/326,529  
Filed: June 4, 1999  
Title: TRANSCRIPTION FACTORS RELATED TO TFIIA  
Current Examiner: Richard Hutson  
Art Unit No.: 1652

Express Mail No.: **EL152182301US**  
Date of Deposit: October 30, 2000  
I certify that the accompanying paper is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to BOX: PATENT APPLICATION, Commissioner for Patents, Washington, DC 20231



Linda L. Gibson

**PRELIMINARY AMENDMENT**

Commissioner for Patents  
Washington, D.C. 20231

Sir:

Please find below Applicant's Preliminary Amendment that accompanies the request for filing a Divisional Application under 37 CFR 1.53(b) filed herewith. Please amend the above-identified patent application as follows:

**IN THE SPECIFICATION**

On page 2, please add the following reference to the co-pending application:

"This application is a Division Application of co-pending United States Patent Application Serial Number 09/326,529, filed on June 4, 1999."

**IN THE CLAIMS:**

Cancel in this Application original claims 1 and 3-32 of co-pending application before calculating the filing fee.

**Please add new claims 71-85:**

71. (New Claim) A purified ALF protein encoded by an oligonucleotide comprising a nucleic acid sequence substantially homologous to the coding strand of the gene sequence set forth in SEQ ID NO:1.

72. (New Claim) The purified protein of claim 71, wherein said protein is a human ALF protein.

73. (New Claim) The purified protein of claim 71, wherein said protein is the human ALF protein of SEQ ID NO.:2.

74. (New Claim) A fusion protein comprising a portion of the ALF protein of claim 71 and a non-ALF protein sequence.

75. (New Claim) The fusion protein of claim 74, wherein said ALF protein is a human ALF protein and said non-ALF protein is SALLF.

76. (New Claim) The fusion protein of claim 74, wherein said non-ALF protein is a transcriptional factor.

77. (New Claim) The fusion protein of claim 74, wherein said non-ALF protein is a human transcriptional factor.

78. (New Claim) The fusion protein of claim 74, wherein said SALLF protein is the human ALF protein of SEQ ID NO.:4.

79. (New Claim) A purified ALF protein encoded by an oligonucleotide comprising a nucleic acid sequence substantially homologous to the coding strand of the gene sequence set forth in SEQ ID NO:3.

80. (New Claim) The purified protein of claim 79, wherein said protein is a human SALF protein.

81. (New Claim) The purified protein of claim 79, wherein said protein is the human ALF protein of SEQ ID NO.:4.

82. (New Claim) A fusion protein comprising a portion of the SALF protein of claim 77 and a non-SALF protein sequence.

83. (New Claim) The fusion protein of claim 82, wherein said non-ALF protein is a transcriptional factor.

84. (New Claim) The fusion protein of claim 82, wherein said non-ALF protein is a human transcriptional factor.

85. (New Claim) The fusion protein of claim 82, wherein said SALF protein is the human ALF protein of SEQ ID NO.:4.

#### COMMENTS

Should any fees under 37 C.F.R. §§ 1.16 to 1.18 be required for any reason relating to the enclosed materials, or should an overpayment be included herein, the Commissioner is authorized to deduct or credit said fees from or to Gardere & Wynne, L.L.P. Deposit Account No. 07-0153.

An action on the merits of all of the claims and a Notice of Allowance thereof are respectfully requested. The Examiner is invited to telephone the undersigned at the telephone number listed below if he or she has any questions or suggested amendments to the claims.

Dated this 30<sup>th</sup> day of October, 2000.

Respectfully submitted:



Edwin S. Flores  
Attorney for Applicant  
Registration No. 38,453  
Gardere & Wynne, L.L.P.  
1601 Elm Street, Suite 3000  
Dallas, Texas 75201-4767  
(214) 999-3000 - Tel  
(214) 999-4667 - Fax

DALLAS 933475v1

## **TRANSCRIPTION FACTORS RELATED TO TFIIA**

### **TECHNICAL FIELD OF THE INVENTION**

The present invention relates in general to the field of recombinant nucleic acids, polypeptides and other derived materials and, more particularly, to the identification, isolation and characterization of human transcription factors that are involved in the expression of human genes.

## BACKGROUND OF THE INVENTION

Without limiting the scope of the invention, its background is described in connection with the isolation, characterization and use of human transcription factors that are expressed throughout the organism, as an example.

5       Unlike the nucleic acid polymerases of prokaryotes, purified RNA polymerase II from eukaryotes initiates transcription very poorly and essentially at random. One key difference between prokaryotic and eukaryotic polymerases is the need for accessory factors that provide for the accurate initiation of transcription. These factors are referred to as the "general" or "basal" transcription factors, in that they are required, in addition to RNA  
10      polymerase II, for the transcription of all eukaryotic protein coding genes. As such, the general transcription factors are expected to be active, or at least present, in all or most tissues. One such general factor is called transcription factor IID (TFIID) and is responsible in large part for promoter recognition. Other general transcription factors include TFIIA, TFIIB, TFIIE, TFIIF and TFIH.

15       Appropriate levels of gene- and tissue-specific transcription is achieved by another set of factors called activator proteins. These factors are often composed of two domains, a sequence-specific DNA recognition domain and an activation domain. When bound to DNA, the activation domain facilitates the formation and function of a preinitiation complex that consists of the general transcription factors and RNA polymerase II. In this way it is possible to direct the selective transcription of genes in an appropriately regulated fashion.  
20

25       The structure of a typical promoter for a eukaryotic gene consists of two general regions. The core promoter is located at or near the actual site of transcription initiation and often includes a TATA sequence element located at about 30 base pairs upstream of the initiation site. The other regions are defined as sequence elements which are recognized by activator proteins. These are often located at various distances further upstream, but may be also be located downstream relative to the core promoter of the gene being regulated. Interactions between bound regulatory factors and the preinitiation complex are responsible for the precisely regulated transcription of each individual gene.

TFIIA is an essential general transcription factor and the purified factor from higher eukaryotes consists of three subunits, designated alpha (35 kD), beta (19 kD) and gamma (12 kD). In humans, the alpha and beta subunits are encoded by DNA sequences present in the TFIIA $\alpha/\beta$  cDNA, sometimes referred to as the 'large' subunit cDNA. These two subunits are post-translationally processed from a large 55 kD product of TFIIA $\alpha/\beta$ . The gamma subunit is encoded by DNA sequences present in the TFIIA $\gamma$  cDNA, sometimes referred to as the 'small' subunit cDNA. This sequence is the subject of United States Letters Patent No. 5,562,117 issued to Moore and Rosen. TFIIA has multiple roles in transcription initiation by RNA polymerase II, including an ability to stabilize TBP-TATA element interactions, displace TBP-associated repressors and serve as a cofactor during the processes of transcription activation.

Most of the known human general transcription factors appear to be generally required in all tissues for gene expression by RNA polymerase II. Thus, these factors will be important as markers to evaluate disease states which may arise from inappropriately regulated gene expression and as pharmacological reagents and/or targets with which to modulate patterns of gene expression. Similarly, overexpression via gene therapy or other means should have broad effects on the expression of many or all cellular genes. In contrast, mutations in the genes for activator proteins, which are normally observed to control expression of a select set of genes, often in a tissue or developmentally restricted pattern, typically result in specific defects. Likewise, overexpression of activator proteins only affects expression of cellular genes which contain cognate recognition sequences.

Testis has important endocrine (hormonal) functions and is the site for the production of haploid spermatozoa from undifferentiated stem cells, a process called spermatogenesis. Mutations in some specialized transcriptional activator proteins, such as A-myb and CREM, cause male infertility and show defects in spermatogenesis. The identification of tissue-specific human general transcription factor would bridge an important gap between the generality for general transcription factor function and the specificity of gene-specific transcriptional activator protein function. If such factors were testis-specific, they would be expected to regulate patterns of gene expression that are important in the endocrine,

spermatogenic and other functions of this organ. The present invention satisfies a need in the art for new compositions for polynucleotide sequences and encoded polypeptide products, immunological reagents and other derived materials in terms of providing unique reagents for the detection of defects in testis function such as idiopathic male infertility or other syndromes, for detection of dysfunctional patterns of gene expression and as reagents that can modulate gene expression.

SUMMARY OF THE INVENTION

The present invention includes DNA sequences that encode two structurally distinct isoforms of the human general transcription factor TFIIA  $\alpha/\beta$ . One of these sequences is denoted as ALF, for TFIIA  $\alpha/\beta$ -like factor, which is expressed predominantly in human testis. The second sequence contains ALF connected to a unique upstream sequence and is denoted as SALF, for Stoned B/TFIIA  $\alpha/\beta$ -like factor. The present invention is also direct to recombinant polypeptide products and other derived materials. The uses of the invention include, but are not necessarily limited to, the propagation and preparation of the ALF and SALF DNA, RNA and recombinant proteins, and use of these materials as reagents and markers to detect and/or modify the function of eukaryotic cells in normal and disease states.

The present invention may be used in the detection of the endogenous ALF and SALF RNAs in eukaryotic cells using hybridization, polymerase chain reactions, immunological analysis and other methods. The invention may also be used along with the endogenous ALF and SALF DNAs, RNAs and proteins as specific *in vivo* pharmacological targets to artificially modulate the expression of eukaryotic genes. Furthermore, the ALF, SALF and the variable carboxyl terminal end may be introduced in a normal or modified versions of the ALF and SALF genes for expression in eukaryotic cells in order to replace or augment endogenous transcription factor activities (gene therapy). The present invention may also be used as testis-specific antigens for contraceptive vaccine development.

The present invention, in a general and overall sense, concerns the isolation and characterization of a novel transcriptional factor gene, ALF and carboxy terminal variable region. One embodiment of the present invention is a purified nucleic acid segment that encodes a protein having an amino acid sequence as shown in Figure 2, in accordance with SEQ ID NO.:2. Another embodiment of the present invention is a purified nucleic acid segment, further defined as including a nucleotide sequence in accordance with SEQ ID NO.:1.

The present invention also concerns the isolation and characterization of a novel transcriptional factor gene, SALF and a carboxy terminal variable region. One embodiment of the present invention is a purified nucleic acid segment that encodes a protein having an

amino acid sequence as shown in Figure 3, in accordance with SEQ ID NO.:4. Another embodiment of the present invention is a purified nucleic acid segment, further defined as including a nucleotide sequence in accordance with SEQ ID NO.:3. The 3' variable region that ALF and SALF have in common is encoded by the nucleic acid segment in accordance with SEQ ID NO.:5 and expressed as an amino acid sequence as shown in SEQ ID NO.:6.

In one embodiment the purified nucleic acid segment includes the nucleotide sequence of SEQ ID NOS.:1, 3 and 5. As used herein, the term "nucleic acid segment" and "DNA segment" are used interchangeably and refer to a DNA molecule that has been isolated free of total genomic DNA of a particular species. Therefore, a "purified" DNA or nucleic acid segment as used herein, refers to a DNA segment that includes novel transcriptional factor genes, ALF, SALF and a carboxy terminal variable coding sequence, yet is isolated away from, or purified free from, total genomic DNA, for example, total cDNA or human genomic DNA. Included within the term "DNA segment", are DNA segments and smaller fragments of such segments and recombinant vectors, including, for example, plasmids, cosmids, phage, viruses and the like.

Similarly, a DNA segment encoding an isolated or purified novel transcriptional factor genes, ALF, SALF and a carboxy terminal variable coding sequence, gene refers to a DNA segment including ALF, SALF and a carboxy terminal variable coding sequence isolated substantially away from other naturally occurring genes or protein encoding sequences. In this respect, the term "gene" is used for simplicity to refer to a functional protein, polypeptide or peptide encoding unit. As will be understood by those in the art, this functional term includes both genomic sequences, cDNA sequences or combinations thereof. "Isolated substantially away from other coding sequences" means that the gene of interest, in this case ALF, SALF and a carboxy terminal variable coding sequence, forms the significant part of the coding region of the DNA segment. Of course, this refers to the DNA segment as originally isolated and does not exclude genes or coding regions later added by the hand of man to the segment.

In particular embodiments, the invention concerns isolated DNA segments and recombinant vectors incorporating DNA sequences that encode novel transcriptional factor

genes, ALF, SALF and a carboxy terminal variable coding sequence genes, and that include within the amino acid sequence an amino acid sequence in accordance with SEQ ID NO.:2. Moreover, in other particular embodiments, the invention concerns isolated DNA segments and recombinant vectors incorporating DNA sequences that encode a gene which includes within its amino acid sequence the amino acid sequence of a ALF, SALF and a carboxy terminal variable coding sequence

Another embodiment of the present invention is a purified nucleic acid segment that encodes proteins in accordance with SEQ ID NOS.:2, 4 and 6, further defined as a recombinant vectors. As used herein the term, "recombinant vector", refers to a vector that has been modified to contain a nucleic acid segment that encodes ALF, SALF, or the carboxy terminal variable coding sequence protein, or a fragment thereof. The recombinant vector may be further defined as an expression vector that includes a promoter operatively linked to the ALF, SALF, or the ALF/SALF variants having the carboxy terminal variable coding sequence encoding a nucleic acid segment.

A further embodiment of the present invention is a host cell, made recombinant with a recombinant vector including ALF, or SALF, and if present, a carboxy terminal variable coding sequence. The recombinant host cell may be a prokaryotic cell. In a one embodiment, the recombinant host cell is a eukaryotic cell. As used herein, the term "engineered" or "recombinant" cell is intended to refer to a cell into which a recombinant gene, such as a gene encoding ALF, SALF, or the carboxy terminal variable coding sequence, has been introduced. Therefore, engineered cells are distinguishable from naturally occurring cells which do not contain a recombinantly introduced gene. Engineered cells are thus cells having a gene or genes introduced through the hand of man. Recombinantly introduced genes will either be in the form of a cDNA, a copy of a genomic gene, or will include genes positioned adjacent to a promoter not naturally associated with the particular introduced gene.

It may be more convenient, however, to employ as the recombinant gene a cDNA version of the gene. One advantage of working with cDNAs is that the size of the gene is generally smaller and more readily employed to introduce into or "transfect" the targeted cell

than will a genomic gene; typically an order of magnitude larger than cDNA gene. Alternatively, a genomic version of a particular gene may be used where desired.

In certain embodiments, the invention concerns isolated DNA segments and recombinant vectors that encode a protein or peptide which includes within its amino acid sequence an amino acid sequence essentially as set forth in SEQ ID NOS.:2, 4 or 6. Naturally, where the DNA segment or vector encodes a full length ALF or SALF protein, or is intended for use in expressing the sequences will be as essentially as set forth in SEQ ID NOS.:2, 4 and 6.

The term "a sequence essentially as set forth in SEQ ID NO.:2" means that the sequence substantially corresponds to a portion of SEQ ID NO.:2 and has relatively few amino acids which are not identical to, or a biologically functional equivalent of, the amino acids of SEQ ID NO.:2. Likewise the phrase is equally applied to SEQ ID NOS.: 4 and 6. The term "biologically functional equivalent" is well understood in the art and is further defined in detail herein as a gene having a sequence essentially as set forth in SEQ ID NOS.:2, 4 or 6, and that is associated with RNA transcription. Accordingly, sequences that have between about 70% and about 80%; or between about 81% and about 90%; or even between about 91% and about 99%; of amino acids that are identical or functionally equivalent to the amino acids of SEQ ID NOS.:2, 4 or 6.

In certain other embodiments, the invention concerns isolated DNA segments and recombinant vectors that include within their sequence a nucleic acid sequence essentially as set forth in SEQ ID NOS.:1, 3 or 5. The term "essentially as set forth in SEQ ID NO.: 1," is used in the same sense as described above and means that the nucleic acid sequence substantially corresponds to a portion of SEQ ID NO.:1, and has relatively few codons that are not identical, or functionally equivalent, to the codons of SEQ ID NO.:1. Likewise the phrase is equally applied to SEQ ID NOS.: 3 and 5. The functionally equivalent codons are known in the art.

It will also be understood that amino acid and nucleic acid sequences may include additional residues, such as additional N- or C-terminal amino acids or 5' or 3' sequences, and yet still be essentially as set forth in one of the sequences disclosed herein, so long as the

sequence meets the criteria set forth above, including the maintenance of biological protein activity where protein expression is concerned. The addition of terminal sequences particularly applies to nucleic acid sequences that may, for example, include various non-coding sequences flanking either of the 5' or 3' portions of the coding region or may include various internal sequences, i.e., introns, which are known to occur within genes.

Excepting intronic or flanking regions, and allowing for the degeneracy of the genetic code, sequences that have between about 70% and about 80%; or between about 80% and about 90%; or between about 90% and about 99%; of nucleotides that are identical to the nucleotides of SEQ ID NOS.:1, 3 or 5 will be sequences that are "essentially as" the respective SEQ ID NOS. Sequences that are essentially the same as those set forth in SEQ ID NOS.:1, 3 or 5 may also be functionally defined as sequences that are capable of hybridizing to a nucleic acid segment containing the complement of SEQ ID NO.:1 under relatively stringent conditions. Suitable relatively stringent hybridization conditions will be well known to those of skill in the art and are clearly set forth herein, for example conditions for use with southern and northern blot analysis as described herein.

Naturally, the present invention also encompasses DNA segments that are complementary, or essentially complementary, to the sequence set forth in SEQ ID NOS.:1, 3 or 5. The nucleic acid segments of the present invention, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. A nucleic acid fragment of almost any length may be employed, with the total length being limited by the ease of preparation and use in the intended recombinant DNA protocol. For example, nucleic acid fragments may be prepared that include a short stretch complementary to SEQ ID NOS.:1, 3 or 5, such as about 10 to 15 or 20, 30, or 40 or so nucleotides, and which are up to 10,000 or 5,000 base pairs in length, with segments of 3,000 being used in certain cases. DNA segments with total lengths of about 1,000, 500, 200, 100 and about 50 base pairs in length are also useful.

Another embodiment of the present invention is a nucleic acid segment that includes at least a 14-nucleotide long stretch that corresponds to, or is complementary to, the nucleic acid sequence of SEQ ID NOS.:1, 3 or 5. In one embodiment the nucleic acid is further defined as including at least a 20, 30, 50, 100, 200, 500, 1000, or at least a 3824 nucleotide long stretch that corresponds to, or is complementary with, the nucleic acid sequence of SEQ ID NOS.:1, 3 or 5. The nucleic acid segment may be further defined as having the nucleic acid sequence of SEQ ID NOS.:1, 3 or 5.

A related embodiment of the present invention is a nucleic acid segment that includes at least a 14-nucleotide long stretch that corresponds to, or is complementary with, the nucleic acid sequence of SEQ ID NO.:1 or 3, further defined as including a nucleic acid fragment of up to 10,000 base pairs in length. Another embodiment is a nucleic acid fragment including from 14 nucleotides of SEQ ID NO.:1 or 3 up to 5,000, 3,000, 1,000, 500 or 100 base pairs in length.

Naturally, it will also be understood that this invention is not limited to the particular nucleic acid and amino acid sequences of SEQ ID NOS.: 2, 4 and 6. Recombinant vectors and isolated DNA segments may therefore variously include the ALF, SALF and variable region coding regions themselves, coding regions bearing selected alterations or modifications in the basic coding region, or they may encode larger polypeptides that nevertheless include ALF, SALF or variable region-coding segments or may encode biologically functional equivalent proteins or peptides that have variant amino acids sequences.

The DNA segments of the present invention encompass biologically functional equivalent ALF, SALF and variable region peptides. Such sequences may arise as a consequence of codon redundancy and functional equivalency that are known to occur naturally. Alternatively, functionally equivalent proteins or peptides may be created via the application of recombinant DNA technology, where changes in the protein structure may be engineered, based on considerations of the properties of the amino acids being exchanged. Changes designed by man may be introduced through the application of site-directed mutagenesis techniques, e.g., to introduce improvements to the antigenicity of the ALF,

SALF or variable region mutants in order to examine transcriptional activity or determine the presence of ALF, SALF or variable region protein in various cells and tissues at the molecular level.

Another embodiment of the present invention is a purified composition comprising  
5 a polypeptide having an amino acid sequence in accordance with SEQ ID NOS.:2, 4 or 2 or  
4 with 6. The term "purified" as used herein, refers to a transcriptional factor protein  
composition, wherein the ALF, SALF or ALF and SALF having the variable region proteins  
are purified to any degree relative to its naturally-obtainable state, i.e., in this case, relative  
to its purity within a eukaryotic cell extract, or a testis sample. A cell for the isolation of  
10 ALF, SALF or variants thereof is a cell of testicular origin, however, these proteins may also  
be isolated from patient specimens, recombinant cells, tissues, isolated subpopulations of  
tissues, and the like, as will be known to those of skill in the art, in light of the present  
disclosure. Purified ALF, SALF or variants thereof also refer to polypeptides having the  
15 amino acid sequence of SEQ ID NOS.:2, 4, 2 and 6 or 4 and 6, free from the environment  
in which it may naturally occur. One may also prepare fusion proteins and peptides, e.g.,  
where the ALF, SALF or variable portion coding regions are aligned within the same  
expression unit with other proteins or peptides having desired functions, such as for  
purification or immunodetection purposes (e.g., proteins that may be purified by affinity  
20 chromatography and enzyme label coding regions, respectively).

Turning to the expression of ALF, SALF and variable genes whether from cDNA or  
genomic DNA, protein may be prepared using an expression system to make recombinant  
preparations of ALF, SALF and variable genes proteins. The engineering of DNA  
segment(s) for expression in a prokaryotic or eukaryotic system may be performed by  
techniques generally known to those of skill in recombinant expression. For example, ALF,  
25 SALF and variable genes-GST (glutathione-S-transferase) fusion proteins are a convenient  
means of producing protein in a bacterial expression. Virtually any expression system may  
be employed in the expression of ALF, SALF and variable gene products. Eukaryotic  
expression systems, however, may also be used.

Transformation of host cells with DNA segments encoding ALF, SALF and variable genes also provides a convenient means for obtaining a protein for ALF, SALF and ALF or SALF including the variable portions. Complementary DNA (cDNA), genomic sequences and combinations thereof, are suitable for eukaryotic expression, as the host cell will, of course, process the genomic transcripts to yield functional mRNA for translation into protein.

Another embodiment is a method of preparing a protein composition comprising growing recombinant host cell comprising a vector that encodes a protein that includes an amino acid sequence in accordance with SEQ ID NOS.:2, 4 or 6, under conditions permitting nucleic acid expression and protein production followed by recovering the protein so produced. The host cell, conditions permitting nucleic acid expression, protein production and recovery, will be known to those of skill in the art, in light of the present disclosure of the ALF, SALF and variable region genes.

BRIEF DESCRIPTION OF THE DRAWINGS

For a more complete understanding of the features and advantages of the present invention, reference is now made to the detailed description of the invention along with the accompanying figures which corresponding numerals in the different figures refer to the corresponding parts and in which:

FIG. 1 depicts the cDNA sequence of ALF (SEQ ID NO. 1);

FIG. 2 depicts the corresponding deduced amino acid sequence of ALF (SEQ ID NO. 2), standard one-letter abbreviations for amino acids is used;

FIG 3 depicts the cDNA sequence of SALF (SEQ ID NO. 3);

FIG 4 depicts the corresponding deduced amino acid sequence of ALF (SEQ ID NO. 4), standard one-letter abbreviations for amino acids is used;

FIG 5 depicts the cDNA sequence of an alternative 3'-coding and untranslated region for both ALF and SALF (SEQ ID NO 5);

FIG 6 depicts the corresponding deduced amino acid sequence of ALF (SEQ ID NO. 6), standard one-letter abbreviations for amino acids is used;

FIG 7A and 7B depicts the schematic structures of the ALF and SALF cDNA sequences. A, The SALF cDNA sequence includes TFIIA $\alpha/\beta$ -like sequences and an upstream Stoned  $\beta$ -like region and is shown with selected restriction enzyme sites. B, The ALF cDNA sequence consists of TFIIA $\alpha/\beta$ -like sequences. The 5'-end of ALF was identified using gene-specific primers 2a2-20 and 2a2-22 and library-specific primers AP1 and AP2. The resulting clone, pRACE22, is shown as a single line. A PCR product that spans the entire ALF sequence (pRACE17) was obtained using gene-specific primer 2a2- 17 and library-specific primer AP1. C, PCR products from SALF are amplified and visualized from human placenta and liver cDNA libraries. The primers used are indicated above each lane (2a2-1, 2a2-6 and 2a2-8). Lanes 5 and 6 are control reactions to which no cDNA template was added;

FIG 8A and 8B are schematic diagrams showing alignments of ALF and SALF sequences with related genes. A, ALF contains conserved regions I, III and IV and an internal nonconserved region II. Beneath ALF are diagrams of TFIIA large subunits from

human (hTFIIA $\alpha/\beta$ ), Arabidopsis (aTFIIA-L), Drosophila (dTIIA-L) and yeast (yTOA1). B, A diagram of the N-terminus of SALF is shown, indicating an upstream serine, threonine, a proline-rich domain and a downstream domain that is homologous to Drosophila Stoned  $\beta$  and the clathrin APs  $\mu 1$  (AP47) and  $\mu 2$  (AP50) proteins.

5

FIG. 9 shows human genomic DNA digested with either BglII or EcoRI and hybridized with an ALF probe. The enzymes are indicated above each lane (lanes 1 BglII; lanes 2, EcoRI). The positions of the molecular weight size markers are indicated to the left of each panel.

10

FIG 10A through 10E are Northern blot analysis autoradiographs of ALF, TFIIA $\alpha/\beta$  and TFIIA $\gamma$  transcripts. Poly(A) mRNA from various human tissues were probed with various gene-specific probes. A, ALF; B, 5'-SALF; C, TFIIA $\alpha/\beta$ ; D, TFIIA $\gamma$ ; and E, actin.

15

FIG 11A through 11F are RNA dot blot analyses autoradiographs of ALF, TFIIA $\alpha/\beta$  and TFIIA $\gamma$  transcripts; A single dot blot containing poly(A) mRNA from multiple human tissues is probed with A, ALF, B, 5'-SALF, C, TFIIA $\alpha/\beta$ , D, TFIIA $\gamma$ ; and E, a ubiquitin control. The source of the mRNA for each spot is listed in F, dashes indicate positions that do not contain mRNA.

20

FIG 12A through 12D show expression and functional analysis of ALF and SALF polypeptides. A, Coomassie-stained SDS-PAGE gel shows that the recombinant histidine-tagged ALF protein migrates at 69 kD (lane 2) and that the recombinant rat TFIIA $\alpha/\beta$  and TFIIA $\gamma$  proteins used in these studies migrate at 55 kD and 12 kD, respectively (lanes 3 and 4). B, p69 (ALF) can substitute for (TFIIA $\alpha/\beta$ ) p55 in stabilizing the interaction between TBP and the Adenovirus Major Late promoter TATA element (-40 to -16). Additions to each reaction are listed above each lane. Polyclonal antiserum against human p55 is added to reactions in lanes 8 (2 $\mu$ l), 9 (4 $\mu$ l) and 10 (4 $\mu$ l). C, Addition of p69 (ALF) and p12 (TFIIA $\gamma$ ) restore activity to transcriptionally inactive TFIIA-depleted HeLa nuclear extracts. Control (undepleted) and TFIIA-depleted extracts are indicated by a "C" and "D", respectively. D, A T7-promoter driven SALF construct produces an [35S]-labeled protein of approximately 170 kD in in vitro transcription- translation reactions.

25

## DETAILED DESCRIPTION OF THE INVENTION

While the making and using of various embodiments of the present invention are discussed in detail below, it should be appreciated that the present invention provides many applicable inventive concepts which can be embodied in a wide variety of specific contexts.  
5 The specific embodiments discussed herein are merely illustrative of specific ways to make and use the invention and do not delimit the scope of the invention.

Human transcription factor IIA (TFIIA) is a cellular factor that, together with additional protein components, regulates the expression of human genes. TFIIA is composed of proteins produced from two nucleic acid sequences (genes), one called hTFIIA  $\alpha/\beta$  and the other hTFIIA  $\gamma$  (Patent No. 5,652,117). These factors are expressed in human tissues and are thought to function primarily through interactions with the TATA-binding protein (TBP), a universally-required eukaryotic protein that functions at promoter DNA sequences to regulate gene expression.  
10  
15

In addition to the DNA sequences themselves, the functionality of the recombinant polypeptides produced from the ALF DNA sequences is shown using several biochemical assays. First, in conjunction with a recombinant TFIIA $\gamma$  polypeptide, the recombinant ALF protein overproduced and purified from *E. coli*, can interact with recombinant human TATA binding protein to stabilize interactions on promoter DNA. Second, in conjunction with a recombinant TFIIA $\gamma$  polypeptide, the recombinant ALF polypeptide can restore RNA polymerase II transcription activity to nuclear extracts from human cells that have been depleted of TFIIA. Related, antibody reagents raised against the recombinant ALF polypeptides react with the corresponding overproduced polypeptides suggesting that immunological detection of the endogenous protein(s) will be feasible. These studies substantiate the prediction that the ALF protein has a role in the regulation of human gene expression.  
20  
25

## **DEFINITIONS**

As used throughout the present specification the following abbreviations are used:  
TF, transcription factor; TBP, TATA binding protein; ORF, open reading frame, EST,

expressed sequence tag; kb, kilobase (pairs); UTR, untranslated region; kD, kilodalton; nt, nucleotide; aa, amino acids; bp, base pairs; PCR, polymerase chain reaction; AP, adaptor protein; DTT, dithiothreitol; PMSF, phenylmethylsulfonyl flouride; EDTA, ethylenediaminetetraacetic acid; IPTG, isopropyl  $\beta$ -D-thiogalactoside; AdML, Adenovirus Major Late; ALF, TFIIA $\alpha/\beta$ -like factor; SALF, Stoned B/ TFIIA $\alpha/\beta$ -like factor.

5 Terms defined herein have meanings as commonly understood by a person of ordinary skill in the areas relevant to the present invention. Terms such as "a", "an" and "the" are not intended to refer to only a singular entity, but include the general class of which a specific example may be used for illustration. The terminology herein is used to describe 10 specific embodiments of the invention, but their usage does not limit the invention, except as outlined in the claims.

The term "ALF" (TFIIA $\alpha/\beta$ -like factor) refers to the nucleotides essentially as set forth (SEQ ID NO. 1) or amino acid sequences essentially as set forth (SEQ ID NO 2). The term "SALF" (Stoned B/TFIIA $\alpha/\beta$ -like factor) refers to the nucleotides essentially as set forth (SEQ ID NO.3) or amino acid sequence essentially as set forth (SEQ ID NO. 4). The term "alternative carboxy terminal domain" refers to nucleotide essentially as set forth (SEQ ID NO. 5) and amino acid sequences essentially as set forth (SEQ ID NO. 6) It is to be understood that alternative carboxy terminal sequence is present as an alternative 3'-coding and untranslated region that can be found on some ALF or SALF transcripts or cDNAs and is meant to be included or implied in all references to the term "ALF and SALF", without referring to this alternative sequence explicitly each time.

20 The terms "a sequence essentially as set forth in SEQ ID NO. (#)", "a sequence similar to", "nucleotide sequence" and similar terms, with respect to nucleotides, refers to sequences that substantially correspond to any portion of the sequences in SEQ ID NOS 1, 25 3 and 5. These terms refer to synthetic as well as naturally-derived molecules and includes sequences that possess biologically, immunologically, experimentally, or otherwise functionally equivalent activity, for instance with respect to hybridization by nucleic acid segments, or the ability to encode all or portions of ALF or SALF activities. Naturally, these terms are meant to include information in such a sequence as specified by its linear order.

The terms "a sequence essentially as set forth in SEQ ID NO. (#)", "a sequence similar to", "amino acid sequence" and similar terms, with respect to amino acids, refers to peptides, polypeptides, proteins, fragments, fusions, derivatives and alterations thereof that substantially correspond to any portion of the sequences in SEQ ID NOS. 2, 4 and 6. These terms refer to synthetic as well as naturally-derived molecules and includes sequences that possess biologically, immunologically, experimentally, or otherwise functionally equivalent activities, for instance, segments of amino acids which possess immunological activity as an antigenic determinant. Naturally, these terms are meant to include information in such a sequence as specified by its linear order.

The term "homology" and "similarity" refers to the extent to which two nucleic acids are complementary. The term "gene" is used to refer to a functional protein, polypeptide, or peptide encoding unit. As will be understood by those in the art, this functional term includes both genomic sequences, cDNA sequences, or fragments or combinations thereof, as well as gene products, including those that may have been altered by the hand of man.

The term "biologically equivalent" or "functionally equivalent" are understood in the art and is further defined herein as sequences that are interchangeable or similarly useful with reference to a given property, for instance selectivity of hybridization, or which encode amino acid segments that have similar properties.

The term "vector" refers to DNA or RNA sequences that have been modified to contain a nucleic acid segment that encodes ALF or SALF, or a fragment thereof. The vector may be further defined as one designed to propagate ALF and SALF sequences, or as an expression vector that includes a promoter operatively linked to the ALF or SALF sequences, or one designed to cause such a promoter to be introduced. The vector may exist in a state independent of the host cell chromosome, or may be integrated into the host cell chromosome

The term "host cell" refers to cells that have been engineered to contain nucleic acid segments from ALF or SALF, or altered segments, whether archeal, prokaryotic, or eukaryotic. Thus, engineered, or recombinant cells, are distinguishable from naturally occurring cells that do not contain recombinantly introduced genes through the hand of man.

5        The term "agonist" refers to a molecule that enhances either the strength or the time of an effect of ALF or SALF and encompasses small molecules, proteins, nucleic acids, carbohydrates, lipids, or other compounds. The term "antagonist" refers to a molecule that decreases either the strength or the time of an effect of ALF or SALF and encompasses small molecules, proteins, nucleic acids, carbohydrates, lipids, or other compounds.

10      The term "altered" or "alterations", or "modified" with reference to nucleic acid or polypeptide sequences is meant to include changes such as insertions, deletions, substitutions, fusions with related or unrelated sequences, such as might occur by the hand of man, or those that may occur naturally such as polymorphisms, alleles and other structural types. Alterations encompass genomic DNA and RNA sequences that may differ with respect to their hybridization properties using a given hybridization probe. Alterations of polynucleotide sequences that encode ALF or SALF, or fragments thereof, include those that increase, decrease, or have no effect on functionality. Alterations of polypeptides refer to those that have been changed by recombinant DNA engineering, chemical, or biochemical modifications, such as amino acid derivatives or conjugates, or post-translational modifications.

15      The term "antibody" refers to any of a class of immunological molecules with the capacity to interact specifically with one or more epitopes of ALF or SALF and includes those that have been altered or modified, for example, 'humanized' antibodies.

20      The term "complementary" refers to polynucleotide sequence that may form stable base pairs with another sequence under a given set of conditions. For instance a complementary strand refers to a sequence that is a sequence of nucleotides whose composition is dictated by the Watson-Crick base pairs (A-T, G-C). This term also refers to any polynucleotide sequence that can selectively interact with another under either permissive or stringent hybridization conditions, as known to those skilled in the art and, in this respect, includes sequences of DNA or oligonucleotides, RNA, protein nucleic acid, other nucleic acid derivatives and fragments thereof, which can reliably detect a particular sequence with a specificity that is useful for genetic or diagnostic studies.

The term "modulate" refers to the ability to effect a change in the structure, function, or regulation of ALF or SALF genes or gene products. These include methods for altering ALF or SALF gene activity and protein function.

5 The terms "segment", "fragment", "portion", "part", "region", or "domain", refers to any subsections, regardless of length, of nucleotide or amino acid sequences set forth in this invention that are either whole or which have been divided either naturally or by the hand of man.

10 The term "oligonucleotide probe" or "oligonucleotide primer" refers to a polynucleotide sequence between approximately 6 nucleotides to 70 to 80 nucleotides, but typically between 15-30 nucleotides, that can be used in direct hybridization, including microarray techniques, or in amplification assays to achieve a substantially selective detection of a complementary sequence. Such sequences are used for a wide variety of additional purposes, the basis of which is the relatively selective annealing to a particular target complementary nucleic acid sequence.

15 The term "sample" refers to any biochemical or biological specimen that is typically being analyzed for some property, for instance biological or biochemical activity of ALF or SALF, presence or absence of ALF or SALF protein, nucleic acid, including any alteration in the normal distribution or structure of these genes or encoded products. Samples may include, but are not limited to, cells and cell extracts or extracts thereof, including protein and membrane fractions, chromosomes, genomic DNA, RNA, cDNA and so forth, regardless of the particular state of isolation.

20 The term "stringent hybridization conditions" refers to an setting in which two polynucleotide sequences are hybridized under conditions that favor specific over nonspecific interactions. For instance, stringent hybridization conditions might include combinations of salt, organic reagents, blocking agents, detergents, temperature and so forth that allow for hybridization between highly similar sequences, such as those related by 95% or greater identity over a length of sequence sufficient for stable hybridization. The term "hybridization" also refers to complementary interactions between a polynucleotide sequence and an oligonucleotide, and may be performed under conditions of varying stringency that

would be dictated by length and homology of the oligonucleotide(s) used and would be determined by one skilled in the art without undue experimentation.

The term "permissive hybridization conditions" refers to an setting in which two polynucleotide sequences are hybridized under conditions in that polynucleotide sequences with less similarity, for instance as low as 50-60%, are caused to interact. For instance, permissive hybridization conditions might include combinations of salt, organic reagents, blocking agents, detergents, temperature, and so forth that allow for stable hybridization, but may allow a greater degree of nonspecific, or background, hybridization than would be observed under stringent conditions. Hybridization between less related sequences is also facilitated by the use of nucleotides such as inosine within hybridizing nucleic acid segments.

The term "transformation" refers to any means by which DNA or RNA is caused to enter a recipient, or host, chromosome, cell, or organism.

The term "detection" refers to the ability to selectively detect a particular biological or biochemical compound. Detection of polynucleotide sequences in a sample, or the levels of such sequences in a sample, is often achieved by hybridization with a complementary polynucleotide or oligonucleotide sequence, or by amplification. Detection of polypeptides is often achieved on the basis of immunological recognition with antigen specific antibodies. Detection of biological or biochemical activity is often achieved by assaying a sample for an activity that is possessed by the compound being assayed.

The term "amplification" refers to the production of multiple copies of a given polynucleotide sequence. Amplification can be achieved as the growth of a vector contained within a host cell. Alternatively, the amplification of specific DNA, or reverse transcribed RNA, sequences that lie between two oligonucleotide primers can be achieved through the polymerase chain reaction, as known in the art.

The term "purified" or "isolated" with reference to DNA or other nucleic acid segment, or amino acid segment, refers to a sequence that includes novel transcription factor genes ALF and SALL, yet is isolated substantially away from, or purified substantially away from total genomic DNA, total cDNA, total or poly(A) RNA, total cellular, subcellular, or tissue extract, or other populations of molecules.

The term "ligand" refers to any molecule, whether nucleic acid, amino acid, or other chemical compositions that interacts with ALF or SALF.

## INTRODUCTION

5       The synthesis of accurately-initiated messenger RNA in eukaryotic organisms requires the assembly of RNA polymerase II and the general transcription factors (TFIIA, B, D, E, F and H) at core promoters (1, 2). Human TFIIA is composed of 35 (a), 19 (b) and 12 (g) kD subunits encoded by the hTFIIA $\alpha/\beta$  (3, 4) and hTFIIA $\gamma$  (5-7) (Patent No. 5,652,117) cDNAs and evolutionarily conserved cDNAs have been characterized in yeast ( $\gamma$ TOA1 and  $\gamma$ TOA2) (8) and Drosophila (dTIIA-L and dTIIA-S) (9-11). The human TFIIA subunits are expressed in all or most human tissues and are thought to function primarily through interactions with the TATA-binding protein (TBP), a universally required eukaryotic protein that functions a promoter sequences to facilitate and regulate gene expression.

10      TFIIA has multiple roles in transcription initiation by RNA polymerase II. First, TFIIA stabilizes the TBP-TATA element interaction (14-17). TFIIA also stimulates transcription by displacing TBP-associated repressors such as Dr1/NC2, Dr2/Topo1, HMG1 and DSP1 (18-22) and counteracts the ability of ADI/MOT1, hTAFII172,  $\gamma$ TAFII145 and hTAFII250 to inhibit TBP binding to DNA (23-27). Second, TFIIA serves as a cofactor for the AP-1, Gal4-AH, Zta, VP16, CTF, NTF and Sp1 activators (4-7, 10, 28-31) and for the PC4 and HMG-2 coactivators (32, 33). Third, TFIIA is required for the isomerization and extension of TFIID-promoter contacts (34, 35) and for stabilizing interactions between TFIID and initiator sequences (36).

15      One of the nucleic acid sequences disclosed herein is called SALF (Stoned B/TFIIA $\alpha/\beta$ -like factor). SALF is composed of both Stoned B/clathrin AP-like and TFIIA $\alpha/\beta$ -like sequences. An initial incomplete SALF sequence was recognized in NCBI database queries using TFIIA $\alpha/\beta$  sequence query by the inventor as an expressed sequence tag (EST) DNA sequence (ID 259637) described by the I.M.A.G.E. consortium (37). Characterization and isolation of additional sequences by the inventor has revealed a composite 3,853 bp cDNA sequence (FIG 3) that contains a 114 nucleotide 5'-UTR and a

161 nucleotide 3'-UTR with a poly(A) addition signal and a 29 nucleotide poly(A) tract. The deduced ORF commences with a putative start codon (AAGATGT) that is preceded by an in-frame stop codon 27 nucleotides upstream and predicts a 1,182-residue polypeptide (FIG 4) with a molecular weight of 132 kD and pI of 5.1.

5 Another new DNA sequence described herein is called ALF, for TFIIA $\alpha$ / $\beta$ -like factor. ALF is isolated by PCR from a human testis cDNA library and is described as a 1,617 bp cDNA (FIG 1) that predicts a 478 amino acid polypeptide (FIG 2) with a molecular weight of 52 kD and pI of 4.4. ALF contains a 15 nt UTR, a putative initiation codon (GTCATGG) that conforms to the Kozak consensus (A/G NNATGG) (38) and 17 bp downstream of the ATG that predict six amino acids (ACLNPV) not present in SALF. ALF is expressed predominantly, if not exclusively, in testis (FIG 10A). Two additional new sequences that contain partial ALF-like sequences connected at nucleotide 1,344 to an alternative 261 bp 3'-end (FIG 5) are identified (I.M.A.G.E. Consortium CloneIDs 785133 and 1657721). These sequences predict a C-terminus in which the last 35 amino acids of ALF are replaced with the residues "AFPRRTSFNT" (FIG 6) followed by a stop codon and a 3'-UTR that contains a poly(A) addition signal and a poly(A) tail. PCR analysis has verified that both ALF and SALF cDNAs which contain this alternative 3'-end are present and can be PCR amplified, from human cDNA libraries. Importantly, none of the sequences disclosed herein have been previously reported, except as partial I.M.A.G.E. Consortium ESTs and as products of the inventor's work as currently disclosed and their intact sequences, structures, functions, uses and other characteristics.

10  
15  
20  
25 A schematic comparison of ALF and other TFIIA large subunit sequences from human (3, 4), Drosophila (9), Arabidopsis (Genbank Accession number X98861) and yeast (8) is shown in FIG 8A. These sequences share a common organization consisting of conserved regions I and IV, acidic region III and an internal nonconserved region II. ALF is similar to its human TFIIA $\alpha$ / $\beta$  counterpart in region I (aa 1-54; 67%) and region IV (aa 417-478; 73%) and in the negatively-charged region III (aa ~340-414; 42% D/E residues). In contrast, region II shares no homology with the corresponding region in hTFIIA $\alpha$ / $\beta$  (or other TFIIA large subunits) and is approximately 100 residues longer. The unique

N-terminus of SALF is 711 amino acids in length (FIG 8B) and contains a region between amino acids 44 to 150 that is rich in proline (20%), serine (21%) and threonine (9%) residues. Residues between 275 and 692 display 47% similarity to the Drosophila Stoned B protein (39) and 46% similarity to an uncharacterized Stoned β-like ORF in C. elegans, C27H6.1 (53). The Drosophila stoned locus was first identified as a class of mutations that caused neurological defects such as temperature-sensitive paralysis (41) and it has been suggested that Stoned B functions in membrane trafficking in neurons (39). In addition, residues from 410 to 692 within the Stoned B-homology region are 33% and 37% similar to the mouse μ1 (AP47) and rat μ2(AP50) clathrin APs, respectively (FIG 7B) (42, 43). The μ1 (AP47) and μ2 (AP50) clathrin APs are subunits of the AP-1 and AP-2 complexes associated with the trans-Golgi and plasma membranes, respectively and function in the internalization, sorting and recycling of receptors and other membrane proteins (44, 45). Thus, the N-terminus of SALF is related to a family of proteins involved in membrane trafficking.

#### ALF AND SALF GENES

One aspect of the present invention is the polynucleotide sequences essentially as set forth as SEQ ID NOS. 1, 3 and 5, and in FIGS. 1 and 3, the complement of these sequences, the RNA versions of both DNA strands and the information otherwise contained within the linear sequence of these polynucleotide sequences and fragments thereof. In the case of nucleic acid segments, sequences for use with the present invention are those that have greater than about 50 to 60% homology with any portion of the polynucleotide sequences described herein, sequences that have between about 61% and about 70%; sequences that have between about 71 and about 80%; or between about 81% and about 90%; or between 91% and about 99%; or which contain nucleotides that are identical, functionally equivalent, or functionally irrelevant, with respect to the nucleotides present in SEQ ID NOS 1, 3 and 5 are considered to be essentially similar. Also encompassed within the present invention are nucleic acids that encode polypeptides that are at least 40% identical or similar to the amino acid sequences shown in SEQ ID NOS. 2, 4 and 6, and in FIGS. 2, 4 and 6.

The invention also encompasses other nucleic acids or nucleic acid like molecules that are sufficient in any regard to mimic, substitute for, or interfere with the ALF or SALF polynucleotide sequences or fragments thereof. It will also be understood that the nucleic acid and amino acid sequences may include additional residues, such as additional 5'- or 3'- sequences, and yet still be essentially as set forth in one of the sequences disclosed herein, so long as the sequence meets the criteria set forth, including the maintenance of functionality, or for the purpose of engineering altered functionality with respect to ALF and SALF.

Included within the invention are DNA or RNA segments including oligonucleotides, polynucleotides and fragments thereof, including DNA or RNA or nucleic acid-like sequences of genomic or synthetic origin, single or double stranded. The invention includes nucleic acid molecules, or nucleic acid-like molecules that are able to hybridize to the sequences in SEQ ID NOS. 1, 3 and 5, under stringent or under permissive hybridization conditions, or to the complement of said sequences.

The invention also includes oligonucleotide, or oligonucleotide-like sequences such as phosphothioates, or peptide nucleic acid sequences, that possess sufficient similarity with the sequences disclosed herein such that they are able to stably hybridize to the disclosed sequences, or their complements. Such sequences may be intended as antisense regulators of gene expression, or for the selective amplification or extension of adjoining sequences, for instance by PCR using a given annealing temperature, as would be determined by someone skilled in the art.

In addition to the sequences disclosed here, related sequences in other organisms, or homologs, will be readily identified by hybridization using the present sequences. This will facilitate the development of animal models for understanding disorders related to the overexpression, underexpression, or expression of forms with altered functionality, with respect to ALF, SALF, and similar sequences. Thus, related genes, and related mRNA transcripts, can be identified by one skilled in the art. The invention thus encompasses methods for the use of the disclosed sequences in various screening procedures aimed at isolating such species. For instance, colony or plaque hybridization techniques can be

performed using radiolabeled sequences as a probe to detect complementary sequences in genomic and cDNA libraries.

Hybridization conditions with respect to temperature, formamide and salt concentrations, in such studies are chosen by one skilled in the art and vary with respect to the organism from which sequences are being isolated, and the sequence similarity, or lack thereof, that is expected based on evolutionary distances. Similar techniques will apply to the isolation of the genomic sequences that encode ALF and SALF, as well as those that encode related genes from organisms other than humans. Reference is particularly made to flanking regions, including upstream sequences that encode the core promoter and regulatory regions, as well as downstream regions, introns and intron/exon boundaries. Similar techniques will also apply to the identification of mutant alleles, polymorphisms, deletions, insertions, and so forth, in genomic and cDNA sequences. These may occur within the ALF and SALF sequences themselves, or may occur in regulatory regions, introns, intron/exon boundaries, or may reflect various insertions, partial or whole gene deletions, or substitutions, any of which may affect biological activity of a gene and gene product. In the case of humans, the identification of interindividual genomic differences in the ALF and SALF genes will be useful in diagnostic determinations.

Whole or partial sequences referred to above may also be identified and isolated using techniques that involve annealing of short oligonucleotides to complementary sequences, such as those as might be present in the genomic DNA of a particular organism, or in genomic or cDNA, including expression cDNA, libraries. Thus, PCR is used to obtain DNA sequences homologous to, and which lie between, two primers, usually between 15 to 30 nucleotides which have annealing temperatures typically between 60-80 degrees Celsius may be substantially purified. The choice of primer sequences, annealing conditions (temperature), number of amplification cycles, choice of polymerase, and so forth would be within the knowledge of one skilled in the art. Amplification assays will be generally applicable to the identification of sequences homologous to ALF and SALF, to the identification of flanking genomic or cDNA sequences, to the identification of mutated alleles, and so forth, in a manner that lends itself to rapid diagnostics.

Variations in PCR technology are also relevant, such as reverse transcriptase mediated PCR, in which mRNA or total RNA is reverse transcribed typically with an oligo dT or gene specific primer prior to PCR amplification. Techniques are also available which utilize only one gene-specific primer, together with a linker or adapter primer as may be present in a vector or attached to the ends of the DNAs to be amplified. For instance, the 5 Genome Walker (Clontech) technique allows the isolation of genomic DNA that flanks a given oligonucleotide primer. Thus, the invention provides a method to isolate the testis-specific ALF gene promoter that can be used to drive cell- or tissue-specific expression of unrelated genes. Techniques are also available in which altered oligonucleotides are employed to generate specific mutations, deletions, insertions, or fusions in the disclosed 10 sequences, or fragments thereof, for instance site directed mutagenesis.

Likewise, the current invention provides methods to map particular regions of a chromosome, and to identify and isolate homologous regions in artificial chromosomes, such as YACs, PACs, single chromosome libraries, and so forth. The current invention also 15 provides techniques such as *in situ* hybridization in order to map disease-associated genes or other chromosomal markers, as well as mutations such as polymorphisms, inversion, translocations, deletions, insertions, and the like, which may be associated with particular health conditions.

Naturally, it will be understood that this invention is not limited to the particular 20 nucleic acid sequences presented herein. Recombinant vectors, including for example plasmids, phage, viruses, and other sequences, and isolated DNA or RNA segments may therefore variously include the ALF and Salf sequences or their complements, and coding regions, as well as those that may bear selected alterations or modifications that nevertheless 25 include ALF or Salf segments or may encode biologically or experimentally relevant amino acid sequences. Such sequences may be created by the application of recombinant DNA technology, where changes are engineered based on the consideration of the nucleotides or amino acids being exchanged, deleted, inserted, fused, or otherwise modified.

Likewise, the current invention encompasses sequences that may be naturally present as extensions of, or insertions within, the sequences disclosed herein, including alternative

or longer 5' or 3' mRNA sequences, or intronic and promoter genomic sequences, or allelic or polymorphic versions of a gene. Similarly, natural, artificial, or synthetic fusions of ALF and SALF, and fragments thereof, with unrelated nucleic acids or amino acids such as those that encode epitope tags, binding proteins, marker proteins, and other amino acid sequences are included.

5

### ALF AND SALF PROTEINS AND POLYPEPTIDES

One aspect of the invention is the protein, polypeptide, oligopeptide, or amino acid sequences or fragments thereof, of ALF and SALF, essentially as set forth in SEQ ID NOS. 10 2, 4 and 6. Sequences that have greater than about 40-50% homology with any portion of the amino acid sequences described herein, sequences that have between about 51% and about 60%; sequences that have between about 61% and about 70% sequences that have between about 70 and about 80%; or between about 81% and about 90%; or between 91% and about 99%; or those that contain amino acids that are identical, functionally equivalent, or functionally irrelevant, for instance those specified by conservative, evolutionarily conserved, and degenerate substitutions, with respect to the amino acid sequences presented 15 in SEQ ID NOS 2, 4 and 6 are included. The invention thus applies to ALF and SALF sequences, or fragments thereof, and nucleic acids which encode such polypeptides, such as those of other species. Reference is particularly, but not exclusively, made to the conserved N- (amino acids 1-54) and C-terminal (amino acids 417-478) regions of ALF and SALF, in contrast to similarity throughout the entire length. The invention thus encompasses amino acid sequences, or amino acid-like molecules, that are sufficient in any regard to mimic, substitute for, or interfere with the ALF or SALF amino acid sequences, or fragments thereof.

20 The invention encompasses ALF and SALF amino acid sequences that have been altered in any form, either through the use of recombinant engineering, or through post-translational or chemical modifications, including those that may be produced by natural, biological, artificial, or chemical methods. Naturally, it will be understood that this invention is not limited to the particular amino acid sequences presented herein. Altered amino acid sequences include those which have been created by the application of 25

recombinant technology such that specific residues, regions, or domains have been altered, and which may be functionally identical, or which may possess unique biological or experimental properties with regards to function or interactions with natural and artificial ligands.

5 For instance such modifications may confer longer or shorter half-life, reduced or increased sensitivity to ligands that modify function, ability to detect or purify polypeptides, solubility, and so forth. Alternatively, such sequences may be shorter oligopeptides that possess an antigenic determinant, or property that interferes, or competes, with the function of a larger polypeptide, for instance sequences similar to the functionally important and conserved N- and C-terminal domains, and those that affect interactions between TFIIA 10 subunits and other proteins. Such sequences may be created by the application of recombinant DNA technology, where changes are engineered based on the consideration of the nucleotides or amino acids being exchanged, deleted, inserted, fused, or otherwise modified. Likewise, the current invention encompasses sequences that may be naturally present as extensions of, or insertions within, the sequences disclosed herein, including alternative or longer N- and C-terminal sequences, or alternatively spliced protein isoforms.

15 Production and purification of polypeptides may be achieved in any of a variety of expression systems known to those skilled in the art, including recombinant DNA techniques, genetic recombination, and chemical synthesis. For instance, expression in prokaryotic cells may be achieved by placing protein coding nucleic acid sequences downstream of a promoter, such as T7, T3, lacI, lacZ, trp, or other cellular, viral, or artificially modified promoters including those that may be inducible by IPTG, tetracycline, maltose, and so forth. Such promoters are often provided for in commercially available recombinant DNA vectors such as pRSET ABC, pBluescript, pKK223-3, and others, or are 20 easily constructed to achieve such a purpose, and often include the presence of multiple cloning sites (MCS) to facilitate restriction digestion mediated cloning of full or partial coding fragments. Such vectors typically contain efficient ribosome binding sites, and in some cases transcription termination signals.

Cells for the expression of such proteins are normally *E. coli*, but could include *B. subtilis*, *Streptomyces* or others prokaryotes. The incorporation of such recombinant DNA can be efficiently achieved by calcium chloride transformation, electroporation, and so forth. In the case of *E. coli*, cells typically grow in LB media with an appropriate antibiotic selection, for instance ampicillin, chloramphenicol, tetracycline and so forth in order to retain the recombinant vector, although vectors which integrate into the cellular chromosome are also possible. The promoter of many recombinant expression vectors require induction by an inducer compound, for instance IPTG, to facilitate high levels of transcription initiation and subsequent protein production. In some instances, nucleic acid sequences within the coding region may be altered to suit the codon usage patterns of a given model expression system or organism.

Peptides, oligopeptides and polypeptides may also be produced by chemical synthesis, for instance solid phase techniques, either manually or under automated control such as Applied Biosystems 431 peptide synthesizer (Perkin Elmer). After synthesis, such molecules are often further purified by preparative high performance liquid chromatography. Thus, the invention provides methods for the production of epitopes for antibody production, or the production of small molecules that enhance or interfere with a specific function or interaction of the ALF or SALF polypeptides.

Methods to produce and purify said polypeptides in eukaryotic systems are widely available and understood by those proficient in the art. Cells for such production are known to include yeast and other fungi, *Drosophila* and Sf9 cells, cells of other higher eukaryotic organisms such as HeLa, COS, CHO and others, as well as plant cells. Similarly, expression could be achieved in prokaryotic or eukaryotic extracts that are able to translate RNAs into proteins, such as rabbit reticulocyte lysates.

Vectors for expression in such systems are widely available both commercially or can be prepared. Such vectors typically are driven by promoters derived from cellular or viral genes, such as CMV, HSV, EBV, SV40, Adenovirus, LTRs, vaccinia, baculovirus polyhedrin promoter, CaMV, TMV, Rubisco, and so forth, and could obviously include the promoters for the ALF or SALF genes themselves. Such vectors are often designed be

regulated by the presence of enhancer or other regulatory element sequences. Introduction of such vectors into cells is often achieved by calcium phosphate or DEAE dextran technologies, liposome mediated techniques, electroporation, or viral mediated infection. Maintenance of such vectors may be achieved by selectable marker such as that conferred by HSV thymidine kinase, HGPRTase, herbicide resistance, visible markers, and so forth.

Selection of an appropriate methodology would be within the scope of those skilled in such methodologies, using the current invention, and would include any combination of host cell and vector which can achieve desired production goals. For instance, the ability of a host cell to drive efficient full-length polypeptide production, glycosylation, membrane anchoring, secretion, absence of contaminating mammalian proteins or infectious agents, proteolytic processing, lipid modification, phosphorylation and so forth may dictate the use of baculovirus/insect cell systems, mammalian cells systems, plant cell systems and so on. In the case of in vitro translation extracts, one embodiment is the coupled transcription and translation of a nonreplicable recombinant vector, where translation is often visualized by the incorporation of a radiolabeled amino acid. The system selected may further depend on the desirability of obtaining purified polypeptides for further characterization, on whether the intent is to evaluate the effect of the overexpressed proteins on cellular gene expression, in vivo or in vitro, to identify compounds that enhance or interfere with the function of the overexpressed polypeptides, or other purposes.

For stable, long term expression, integration within the host cell chromosome, or as an autonomously replicating element, may be used. ALF or SALF genes, including defective (knock-out) genes themselves, can also be introduced to produce transgenic animals, for instance rodents, primates, insects, and other organisms. These methods provide an opportunity to develop and study animal models for specific gene defects, or for augmented expression of certain genes. Such techniques include pronuclear microinjection, retrovirus mediated transfer and other viral vectors, gene targeting into embryonic stem cells, homologous or nonhomologous recombination and electroporation. The presence and expression of transgenes may occur in all or some cells of a given organism. Likewise, expression of the transgene may be constitutive or inducible and may occur in all or only

some cell types. Characterization of the introduced transgene, or mutant (knock-out) construct is typically achieved by genomic Southern blotting and/or PCR analysis of genomic DNA, and its expression by RNA-RNA, DNA-RNA, DNA-DNA hybridization such as Northern analysis, or by RT-PCR analysis.

5       The invention also relates to cells which contain such recombinant constructs, where the host cell refers to mammalian, plant, yeast, insect, or other eukaryotic cells, or to prokaryotic, or archae, and vectors that are designed for a given host. Promoter-vector combinations could be chosen by a person skilled in these arts. In some cases, the desired outcome may not be protein, but RNA, and recombinant vectors would include those with 10 inserts present in either forward or reverse orientations.

Many of the vectors and hosts have specific features that facilitate expression or subsequent purification. For instance DNA sequences to be expressed as proteins often appear as fusion with unrelated sequences that encode polyhistidine tags, or HA, FLAG, myc and other epitope tags for immunochemical purification and detection, or phosphorylation sites, or protease recognition sites, or additional protein domains such as glutathione S-transferase (GST), maltose binding protein (MBP), and so forth which facilitate purification. Vectors may also be designed which contain elements for polyadenylation, splicing and termination, such that incorporation of naturally occurring genomic DNA sequences that contain introns and exons can be produced and processed, or such that unrelated introns and other regulatory signals require RNA processing prior to production of mature, translatable RNAs. Proteins produced in the systems described above could be subject to a variety of post-translational modifications, such as glycosylation, phosphorylation, nonspecific or specific proteolysis or processing.

Purification of ALF, SALF, or carboxy terminal variants produces as described above 25 can be achieved by any of several widely available methods. Cells may be subject to freeze-thaw cycles or sonication to achieve disruption, or may be fractionated into subcellular components such as nuclear and cytoplasmic fractions prior to further purification. Purification may be achieved by one or more techniques such as precipitation with salts or organic solvents, ion exchange, hydrophobic interaction, HPLC and FPLC chromatographic

techniques. Affinity chromatographic techniques could include the use of polyclonal or monoclonal antibodies raised against the expressed polypeptide, or antibodies raised against or available for an epitope tag such as HA or FLAG. Similarly, purification can be aided by affinity chromatography using fusions to the desired proteins such as GSH-affinity resin, maltose affinity resin, carbohydrate (lectin) affinity resin or, in a one embodiment, Ni-affinity resin, and so forth. In some instances purification is achieved in the presence of denaturing agents such as urea or guanidine, and subsequent dialysis techniques may be required to restore functionality, if desired.

## 10 ANTIBODIES TO ALF AND SALF PROTEINS

The current invention encompasses antibodies of any class, such as IgA, IgD, IgE, IgG, IgM, and subclasses, including polyclonal, monoclonal, chimeric, single chain, humanized and antibody fragments, including synthetic antibodies as in recombinant antibody expression library, single chain antibodies, anti-idiotype antibodies and other immunological, or binding, factors that recognize one or more epitopes of the ALF and SALF proteins. Such reagents as derived from ALF and SALF provide methods for detection and purification of ALF and SALF polypeptides, including endogenous, recombinant, or synthetic factors, and as a means to affect changes in gene expression or other functions by immunochemically targeting ALF or SALF. Proteins, polypeptides, oligopeptides, or peptides will be suitable for the production of monoclonal and/or polyclonal antibodies against the ALF, SALF, and alternative C-terminal regions, and for the use as standards or controls in assays such as ELISA, RIA, FACS, Western analysis, and so forth.

Antigens used to generate antibody reagents have a length of at least five amino acids, and in some cases 10 or more, up to the length of the full-length protein. Techniques that are used to obtain such reagents are described in, e.g., Harlow, et al., *Antibodies: A laboratory Manual*, Cold Spring Harbor Laboratories, New York (1988). For polyclonal antibodies, animals such as rabbits, mice, rats, goats, and so forth are injected with up to several hundred micrograms of antigen together with an adjuvant such as Freund's, either complete or incomplete, followed by a series of booster injections. Blood (serum) collected at intervals

following injections are tested for antibody titer and specificity. Use of such antibodies is often facilitated by further purification by salt fractionation, antigen affinity chromatography, or other purification methods, to obtain more pure, and thus more specific, antibodies.

Monoclonal antibody production can be achieved by several methods, including the hybridoma technique, the human B-cell hybridoma technique and the EBV-mediated hybridoma technique. Thus, in one embodiment, antibody producing cells from mice spleen are fused with myeloma cells. Hybrids are then subcloned and screened for antibodies with the antigen, in this case ALF, SALF, the alternative C-terminal domain, or related antigenic fragments. Antibodies against related polypeptides which are intended for the same purpose constitute an included methodology for detection.

Humanized antibodies are those in which the antigen recognition region from an antibody with a given specificity that possess domains from human antibodies so as not to be recognized or rejected by humans. Chimeric antibodies are those in which genes for the antigen recognition region and the constant regions are spliced from different organisms, for instance mouse and human.

In another embodiment, ALF or SALF antigens may be used to generate an immunological reaction, as described above, which generates an immunological response that, in turn, causes a biological or developmental effect, such as modulation of gene expression, or which affects a cellular process such as testis function or spermatogenesis.

## **DETECTION OR DIAGNOSIS OF ALF OR SALF GENES, GENE PRODUCTS AND ABNORMALITIES THEREOF.**

One embodiment of the invention is the use of the invention for the detection of DNA and RNA sequences of ALF, SALF and the alternative carboxy terminus. Such efforts might be directed towards evaluating the levels of these polynucleotides, to evaluate whether such sequences are present or absent in given individuals, or to evaluate whether corresponding sequences in given individuals are in some way absent, abnormal, or otherwise altered. Thus, the invention encompasses methods and reagents for the production and use of oligonucleotide probes, or DNA or RNA probes of various lengths, that have sufficient

similarity to ALF and SALF nucleic acids to allow for selective detection. Methods for labeling could employ radioactive nucleotides, e.g., using T4 polynucleotide kinase, DNA polymerases, in vitro synthesis of RNA probes, PCR amplification of labeled DNAs, as well as nonradioactive techniques such as incorporation and detection of fluorescent, chromogenic, chemiluminescent compounds, as well as avidin/biotin based systems.

Other detection methods could include those based on direct hybridization, such as include fluorescent in situ hybridization (FISH), in situ hybridization, DNA "chip", or "microarray" hybridization technology, Southern and Northern hybridization analysis, RNA dot blot hybridization, dipstick, pin, dot blot, in situ PCR, and other techniques. Others methods may be based on annealing between short, typically 15-30 nucleotide, complementary DNA or RNAs followed by enzymatic extension, such as PCR analysis of genomic DNA or cDNA, reverse-transcriptase mediated PCR using RNA. Such oligonucleotides are derived from the sequences in SEQ ID NOS 1, 3, or 5, or their complements, or from flanking 5'- or 3' cDNA or genomic regions, or from introns or alternatively spliced exons, alleles, promoter or enhancer regions, and so forth. Visualization of the results from such methods is accomplished by a number of methods, including light or fluorescent microscopy, autoradiographic detection, or detection based on ethidium bromide stained agarose gels, DNA sequencing, and so forth. These and other techniques would be those available to and recognized by those skilled in the art.

To provide a basis to establish whether gene structure or expression correlates with a given pathology, results between normal and experimental subjects are compared with respect to the sizes of hybridizing bands, nucleic acid sequence differences, quantitative differences in gene copy number and expression. Thus standard values from normal individuals are compared to those from individuals which display a particular set of symptoms to determine if symptomatic individuals fall outside normal deviations. Such detection may be facilitated by ELISA, or microplate, type assays in which a chemical composition may be coupled to a support to facilitate reading of multiple samples rapidly, for instance in an automated format, as judged by some spectrophotometric or colorimetric response. Such techniques could also be used to assess the efficacy of other treatments

related to the function or production of ALF or SALL genes and gene products. The ability to detect corresponding DNA or RNA sequences could be provided for in an appropriately licensed pharmaceutical kit that would contain sequence-specific reagents capable of selective detection. Such detection might be made with whole genomic DNA, for example from blood or other tissues, or from RNA or DNA obtained from cultured cells or sampled tissues, or using chromosomes of cells, as examples.

Comparison among samples derived from given individuals using a given detection methodology, for example differential hybridization, conformational polymorphisms, sequencing, and so forth would be compared to those from normal controls. Such procedures may be directed towards the detection of ALF, SALL, and alternative carboxy terminal sequences themselves, or to corresponding genomic sequences, including the promoter and intronic sequences, or to the lack of these sequences due to deletion. Detection procedures would enable differences in test samples, for example those which may be testis-derived, including those which may be dysfunctional or cancerous, to be compared with normal samples.

In one embodiment of the invention, a polynucleotide sequence derived from those disclosed herein is used as a target in a microarray as a means to identify the presence or absence of expression, the presence or absence of gene mutants, and so forth. For instance, an oligonucleotide, or oligonucleotides, are synthesized (typically between 6 and 70 or 80 nucleotides long) that are computer optimized for minimal secondary structure and minimal likelihood of nonspecific interactions. These oligonucleotides, or mixed sets of oligonucleotides, are then coupled to, or synthesized directly on a support, usually in a grid arrangement, such as nylon, glass, or other membrane, wafer, chip, slide, and so forth. Labeled probes from a biological sample may be either DNA, RNA, or reverse transcribed DNA, are used to hybridize to the microarray. After washing, the retention of the probe to a given oligonucleotide is typically determined by fluorescence analysis in a scanner. These techniques are understood by one of skill in the art and performed in accordance with the appropriate instrumentation.

Alternatively, the invention provides a method for isolating identical or related genes from humans or other organisms that may have similar functionality, for instance by PCR or hybridization analysis of genomic and/or cDNA libraries. The invention also provides a means to identify the corresponding genomic DNA and the corresponding tissue-specific promoter DNA and regulatory sequences.

The present invention provides a means to use the ALF or SALF specific antibodies described above in the form of an appropriately licensed kit or pharmaceutical pack which contains reagents and supplies for detecting the expression of relevant polypeptides in cells and tissues. Such approaches include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), fluorescent activated cell sorting (FACS). Another approach is a Western blot, in which crude protein extracts are separated on SDS-PAGE gels, transferred to nitrocellulose and hybridized with a specific antibody. Hybridization typically involves prior blocking with nonspecific proteins and dilution of antibody to a level that facilitates specific as opposed to nonspecific interactions. Visualization typically occurs by the action of a secondary antibody which is typically coupled with a domain such as horse radish peroxidase or radiolabeled protein A which possesses an activity or property which can be visualized. Samples tested may be identical in size and abundance to those observed in normal samples, or they may be increased or decreased in abundance, or they may possess altered mobilities.

Likewise, the level of ALF and SALF antigen in a given sample may be evaluated through the use of an ELISA assay, in which samples are attached to a support, typically a set of plastic wells and are incubated with specific antibody. After washing, secondary antibodies are added that allow visualization, and indicate levels, of the primary antigen which are present. Controls, or standards, would consist of recombinant antigen present in a series of dilutions, as a standard curve. Plates are then read with a microplate spectrophotometer. Antibodies are also used for immunolocalization of antigenic proteins within tissues and cells, to determine appropriate expression and localization of antigen among various subcellular compartments. Antibodies may be employed in specific immunoprecipitation or immunopurification protocols that are designed to identify and

isolate corresponding cellular antigens, or complexes that may contain such antigens. Differences among cells or tissues that may be causative with respect to abnormal biological function are identified in this way.

The invention also provides a means to identify endogenous levels of ALF and SALF activity. Tissue samples or biopsies are obtained and extracts of nuclear proteins or whole cell extracts are produced by homogenization in a physiological buffer together the presence of protease inhibitors, if necessary. Extracts are then tested directly for activity, for instance as described in the TBP-dependent bandshift assay. Alternatively, further precipitation or chromatographic purification may sometimes be necessary to isolate and differentiate endogenous activities away from crude samples.

## **IDENTIFICATION OF LIGANDS AND INTERACTION PARTNERS OF ALF AND SALF**

The invention provides a method of identifying compounds, whether they be specific chemicals or drugs, polypeptide fragments of ALF, SALF or other proteins, in vitro or in vivo selected oligopeptides isolated by selections such as phage display, or RNA or DNA sequences that with affinity for such proteins and complexes. Such compounds are typically identified based on affinity-based interaction assays, or on their ability to modulate function in simple representative functional assays, for instance, modulation of TBP-TATA interactions as visualized by bandshift assays. Such compounds are typically next tested in an in vitro functional assay such as that described to test their ability to interfere with (antagonists), or facilitate (agonists), the function of RNA polymerase II, either in terms of basal or activated transcription. Such compounds may be tested in tissue-derived cells, or in cell lines for modulation of activity, for instance gene expression. Such compounds are then evaluated for their efficacy as therapeutic agents, particularly with respect to the modulation of gene expression in a tissue or organisms itself.

Thus, such compounds may be added individually, or in pools, to assess whether in the presence of the polypeptides, RNA polymerase II and other required general factors and upstream activators, and they can affect initiation and chain elongation of RNA from

particular promoters, including those which may be tissue- or cell-type specific. For instance, compounds may be tested for their ability to modulate the expression of genes that require ALF or SALF function, as judged by visualization of a reporter such as beta-galactosidase, luciferase, G-free cassette, and so forth. Similarly, the invention provides  
5 a method to identify compounds that could confer could confer an activated phenotype, or a dominant negative phenotype, in terms of the polypeptides themselves. In particular, mutations are constructed according to site-directed mutagenesis techniques or by random mutagenesis techniques that are known to practitioners of the art, and selected and evaluated for functionality. Further, computer modeling of the ALF or SALF polypeptides, either  
10 based on X-ray crystallographic data, or on comparisons with known TFIIA structures provides a method to fit, or design, compounds that may interfere with various functions or interactions of ALF or SALF, for instance subunit- subunit interactions, subunit-TBP interactions, and so forth.

The method also provides for solid phase high throughput screening methods for identifying ligands which interact with ALF or SALF. For instance ALF or SALF, alone or with a TFIIA $\gamma$  subunit, or even with TBP, promoter DNA, or other factors, are linked to a solid support. Compounds to be tested for interaction are co-incubated, washed to remove nonspecific or less specific interactions, and those that remain are of relatively higher affinity. Detection is achieved by any of a number of techniques, including antibody linked markers, radioisotopic counting and so forth, preferably via automated sample handling.  
15

Identification of ligands may also be achieved using multiple rounds of PCR-mediated selection of high affinity nucleic acid ligands, in which interacting ligands are typically separated from non-interacting ligands by gel shift. Likewise, the promoter element sequences that have affinity for ALF or SALF containing complexes, for instance those which contain TFIIA $\gamma$  and even TBP, can be selected for by multiple rounds of PCR amplification starting from either random or TATA-containing oligonucleotides. The derived sequences may be compared to the genomic DNA sequences, such as promoters.  
20

Factors such a proteins, or small peptides that interact with ALF or SALF may be identified by phage display, in which a library of oligopeptides with diverse sequences are

detected by virtue of their ability to be recognized, or selectively retained by, ALF or SALF polypeptides that are labeled, or fixed to solid support. Interacting phage are detected, reamplified and reselected until single isolates, or plaques are obtained. The sequence of the interacting motif is identified by sequencing all or part of the coat protein which has been engineered to contain such sequences.

The invention also provides techniques to identify natural interacting proteins of ALF and SALF. For instance, immunoprecipitation using antibodies specific to ALF, or SALF, or to epitope tags that have been engineered within recombinant ALF or SALF nucleic acids, can be used to remove, or pull-down, ALF and SALF containing complexes from cellular extracts, for instance those derived from testis tissue or testis cell lines. Such techniques can be performed either in solution, or with the antibody linked to a solid support, such as a resin or a well.

Similarly, ALF or SALF with GST-encoding nucleic acids in the form of a recombinant expression vector allow overproduction and purification of the fusion protein. Such a protein may be bound to affinity resin such as S-hexyl glutathione, and cellular or recombinant proteins, or fragments thereof, can be tested for their ability to interact with ALF or SALF. Such an analysis typically involves comparisons of the bound proteins compared to eluted proteins using a resin that is loaded with the fusion protein and a resin that is loaded with the GST domain alone.

The yeast two-hybrid and related systems also provide methods to identify interacting factors. In this method, genetic fusions of DNA binding domains and activation domains are made separately to the target and bait polypeptides and clones that express interacting epitopes are identified based on a transcription activation assay *in vivo*. This may be performed in variety of cell types, including yeast and mammalian cells.

The invention provides a means to identify genes which may be selectively regulated by the ALF or SALF polypeptides. For instance, targets of ALF or SALF function may be identified in cells which are engineered to overexpress, or underexpress the ALF or SALF genes, or altered forms of these proteins. RNA isolated from such cells may be purified and compared to that present in normal cells. For instance, differential display, subtractive

hybridization and microarray techniques are available to evaluate differences in gene expression, or RNAs, present in two or more populations. For use with microarray analysis, gene specific oligonucleotides or segments, typically 6 to 70 or 80 nucleotides are coupled to, or synthesized on, a solid support and mRNA populations (cDNA) are prepared from normal control cells or tissues and from cells or tissues which are either overexpressing ALF or SALF, or which contain knockouts in these genes. After hybridization and washing, the microarrays can be scanned for hybridization, for instance as registered by fluorescence. In this way genes whose expression is unaltered, decreased, or increased in response to the presence or absence of the transcription factors ALF or SALF can be identified.

Likewise, serial analysis of gene expression (SAGE) analysis provides a method in which short sequences derived from cDNAs from two populations of mRNA are quantitatively compared for their frequency of identification. The technique relies on the use of restriction enzymes that cut away from their binding site and the concatamerization of the resulting fragments into a vector for sequencing. Thus, single sequencing runs of each isolate can give data from multiple original cDNAs. Confirmation of the isolates is often achieved by Northern or quantitative or semi-quantitative PCR analysis. Thus, engineered ALF and SALF genes provide a method to identify additional genes that are "downstream" or regulated, by ALF or SALF, which may in turn be targets for detection, diagnosis and intervention with regard to correlating disease conditions.

#### **TREATMENT OF ALF- AND SALF-RELATED DISORDERS**

The invention provides methods for the isolation, detection, diagnosis, development of animal models, or therapeutic protocols applicable to any organism, such as cats, dogs, pigs, cows, horses, rabbits, birds, primates and humans. Thus, nucleic acid sequences, antibodies and other agonistic or antagonistic ligands may be used to inhibit or augment patterns of gene expression by modifying ALF and SALF function, or the function of the ALF or SALF genes themselves. Such compounds could be delivered in various media, including buffered saline or other carriers or solvents dependent on the chemical nature of the compound and the route of delivery. Delivery mechanisms could include intranasal,

subcutaneous, intramuscular, intraperitoneal, intradermal, intravenous, topical, enteral, rectal, intramedullary, intraarterial, sublingual, or other means. Doses would vary depending on the need to alleviate or correct particular symptoms and on the particular agent. Standard measures of effectiveness and toxicity in cell cultures and animals are given by ED<sub>50</sub>, the dose that is therapeutic for 50% of the sample, and LD<sub>50</sub>, the dose that is lethal to 50% of the sample. Administration of polypeptides, drugs and other therapeutic compounds would be apparent to those skilled in the art with respect to the present teachings, and in accordance with licensing and regulatory requirements. Normal dosages range from 0.1 ug to 100 mg, up to a total dose of 1g, depending on the particular formulation, delivery route, patient sensitivity, patient history, clearance rates, half-life, and other considerations as would be generally available to one skilled in the art. Compositions with high therapeutic indices will generally be used. Additional details are available in such references as Remington's Pharmaceutical Sciences. Initial estimates as to effective concentrations may be determined using cells grown in culture, or in animal models. The invention also provides for the use of proteins, agonists, antagonists, nucleic acid sequences and vectors administered in conjunction with other therapeutic agents, according to accepted usages by those skilled in the art.

Concentrations for delivery of nucleic acids, polypeptides, antibodies and transgenics will be specific to particular cells, conditions, etc. For instance, the invention provides methods for the use of oligonucleotides, phosphorothioate oligonucleotides, peptide nucleic acids and other nucleic acid-like, or nucleic acid binding molecules that could be administered in a manner and amount designed to treat the specific indications using antisense therapy. Such sequences may contain additional, or other modifications, such as methylation, acetylation, thiolation of normal bases, as well as the use of unconventional bases in order to enhance stability. Such complementary sequences are generally targeted along the coding or control regions of ALF or SALF, so as to control transcription or translation and may be supplied as a drug, or as an antisense transcript, or other functional molecule such as a ribozyme, derived from from a transgene.

Similarly, pharmaceutical intervention would also be applicable to specific antibodies and to other compounds that target ALF and SALF function. Therapeutic antibodies that recognize ALF or SALF may be used to directly target, and inactivate, ALF and SALF polypeptides through antibody-epitope interactions, possibly when coupled with compounds that facilitate cellular entry. Likewise, such antibodies may serve as a carrier or targeting molecule by which other therapeutic compounds might be brought to cells which harbor ALF or SALF.

### ENGINEERED ORGANISMS AND ANIMAL MODELS

The present invention provides methods for the production of engineered cells and organisms such as rodents, but which may include humans or other organisms in need of therapy, that express ALF or SALF or altered forms in the form of a recombinantly introduced gene (transgene). For instance, cells can be engineered with polynucleotide sequences so that expression of active or dominant negative ALF or SALF polypeptides are produced, for instance to compensate for the loss, or overactivity, of such polypeptides in a patient. The current invention also provides methods for ex vivo gene therapy, in which recombinant vectors are introduced into stem cells, or other accessible cell population, and, after engineering and propagation, are transplanted back into the patient. Such engineered constructs might be designed to produce RNA that will be translated into ALF or SALF proteins, or altered versions thereof. Alternatively, such constructs might be designed to produce antisense RNA designed to inhibit transcription or translation, or to produce ribozymes that target ALF and SALF RNAs for degradation. Alternatively, the current invention also provides methods by which deletions, or knock-outs of the ALF or SALF genes can be produced, in order to establish animal models for pathological conditions that result from the absence of these genes.

These methods are known to those in the art, and may include stable integration of DNA sequences by recombination, adenoviral, retroviral and other means which are intended to introduce and propagate sequences in engineered cells, by techniques such as pronuclear microinjection, liposome mediated uptake, electroporation of embryos, homologous

(targeted) recombination, and so forth. One embodiment is receptor-mediated gene transfer, whereby the transgene is coupled with a ligand via polylysine, where the ligand is some molecule that interacts selectively with surface molecules, or receptors, on a selective cell population such as those that might be present in testis, or an antibody that has specificity for

5

a cell-specific surface marker.

Depending on the method chosen, the introduced gene might replicate autonomously as part of a vector, or may integrate a specific or random sites. Such a gene may be engineered so as to contain regulatory sequences that drive expression in a constitutive, inducible, tissue- or cell-cycle specific, or other manner, as desired. Another embodiment 10 of the invention is the use of homologous recombination targeting vector that contains a recombinant ALF or SALF gene that has been engineered to be nonfunctional. Such a construct can be used replace the endogenous gene by homologous recombination, for instance, in embryonic stem ES cells from mice, followed by selection, implantation and development of the modified cells into adult organisms which contain targeted defects in 15 ALF or SALF genes.

10

15

20

25

The presence of the transgene, or knockout construct, may be established by genomic blotting or genomic PCR, and its expression, or lack thereof, by Northern blotting or RT-PCR, or other hybridization technologies. Transgenes that express ALF or SALF proteins, or altered forms, can be further detected using antibodies specific for the expressed protein. Once transgenic founder animals are produced, they are bred to produce colonies of animals with particular genotypes, including inbreeding and outbreeding to develop homo- and heterozygous animals, with respect to the transgene, in different backgrounds. Cells derived from such animals may also be isolated and propagated for study. Further examination at the organismal, tissue, cellular, subcellular and biochemical levels will establish the biological effects of organisms that contain engineered ALF or SALF genes. Such organisms and cells will also provide assay systems with which to identify agonists and antagonists that may compensate for observed defects, and which may perform similar functions in other organisms, including humans.

The invention may be better understood with reference to the following examples. These examples, however, should not be taken to limit the scope of the invention in any way.

## EXAMPLE 1

### CLOTHING AND DETECTION OF ALF AND SALF

The 5'-end of SALF was amplified by PCR (40 cycles) using 4 µl of the human placental cDNA library (Clontech) with primer 2a2-6 (5'-AGTAACCCGAATGCTTAA- 3') (SEQ ID NO.: 8) and a commercially available library-specific adapter primer AP1 (Clontech). The resulting products were reamplified (35 cycles) with primer 2a2-8 (5'ATGCTAGCTGAACCACTG-3') (SEQ ID NO.: 9) and a commercially available nested library-specific adaptor primer AP2 (Clontech) used to obtain a 2,930 bp product, which was subcloned into the pCRII cloning vector (Invitrogen) to form pRACE4 (FIG 7A). Sequence analysis of this and EST ID 256637 constitute SEQ ID NO. 3. Human SALF cDNAs were identified by PCR amplification (35 cycles) of 1.1 and 0.9 kb products from human placenta, liver and testis "Marathon" cDNA libraries (Clontech) using 25 pmol of the upstream primer 2a2-1 (5'- AGAAATTCCCTCTGATTG-3') (SEQ ID NO.: 7) and the downstream primers 2a2-6 and 2a2-8. The 1.1 and 0.9 kb products derived from the liver cDNA library were subcloned into pGEM-T Easy (Promega). Sequence analysis of the liver-derived products shows that they are identical to those present in both SEQ ID NO. 3. These products are diagrammed in FIG 7A, and shown in FIG 7C.

ALF sequences were isolated by PCR (35 cycles) using 4 µl of the human testis cDNA library (Clontech) with the gene-specific primer 2a2-20 (5'-CCAGAAGGTAGAATTGCGGGTTGCTGTAGC-3') (SEQ ID NO. 12) and primer AP1 (Clontech), and reamplified with 2a2-22 (5'-GGAGTTGAAGTGCCCAGGTCTGCTGTGG-3') (SEQ ID NO.: 19) and primer AP2 (Clontech). The 369 bp amplification product is subcloned into pGEM-T Easy (Promega) to form pRACE22. The resulting clone (pRACE22; FIG 7B) contains 298 bp that are identical to SALF and a 35 bp 5'-end that is unique to SEQ ID NO. 1. A full-length ALF PCR product is amplified (35 cycles) from 4 ml of the testis cDNA library (Clontech) using

primer 2a2-17 (5'- GGTGCTGTCATGGCCTGCCTCAACCCGG-3') (SEQ ID NO. 13), located within the unique 5'-end of ALF, and primer AP1 (Clontech). The resulting 1.7 bp fragment is subcloned into pGEM-T Easy to form pRACE17 (FIG 7B). The sequence of the resulting clone is identical to the composite SALF sequence except for its unique 5'-end and a longer poly(A) tail (~90 nucleotides) which begins four nucleotides downstream of the poly(A) tail in SALF. The composite sequences of pRACE17 and pRACE22 are SEQ ID NO. 1.

## EXAMPLE 2

### DETECTION OF GENOMIC DNA SEQUENCES THAT ENCODE ALF.

Genomic DNA (10 µg) from HeLa cells is digested with the indicated restriction enzymes (BglII and EcoRI), electrophoresed on 0.7% agarose gels using 1X TBE buffer, and transferred overnight to nitrocellulose membranes (Schleicher and Schuell). Hybridization is performed under stringent conditions at 42° C in 50% formamide, or at 68° C, in hybridization buffer (6X SSC, 0.5% SDS, 5X Denhardt's solution, and 100 µg/ml salmon sperm DNA). The probe is a full-length NdeI-BamHI ALF fragment contained within the vector construct pRSET-ALF. The blot is washed at 65° C in 0.1X SSC and 0.5% SDS, and exposed at -80° C to XAR-5 film (Kodak). Hybridization with the ALF probe revealed bands of 8.6, 6.9, 5.0, and 1.0 kb (BglII; lane 1), or 11.5, 8.4, 6.0, and 4.5 kb (EcoRI; lane 2). These results (FIG 9) show that sequences complementary to ALF are present and detectable in human genomic DNA.

## EXAMPLE 3

### EXPRESSION OF ALF AND OTHER HUMAN TFIIA SUBUNIT mRNAs

Northern blots containing 2 ug of poly(A) mRNA from 16 human tissues are obtained from Clontech. Gene-specific probes for hybridization are as follows: ALF, a 621 bp NcoI-KpnI fragment or an 899 bp HincII-BglII fragment from region II (FIG 7B); 5'-SALF, a 1,002 bp EcoRI-EcoRI fragment from pRACE4 containing the 5'-UTR and nucleotides

encoding the first 282 residues (FIG 7A); hTFIIA $\alpha/\beta$ , a full-length 1.1 kb EcoRI-EcoRI fragment from lambda11 or a 282 bp HaeIII-Hae III fragment from region II; hTFIIA $\gamma$ , a full-length 355 bp NdeI-BamHI fragment or a 262 bp NdeI-EcoRI fragment from pRSEThp12; and actin controls (Clontech). DNA fragments were typically labeled with [ $\alpha$ 32P]-dCTP using Ready-to-Go DNA Labeling Beads (Pharmacia) and purified over NICK columns (Pharmacia). Northern blots are hybridized for 1 hour in ExpressHyb solution (Clontech) and washed at 68° C for 1 hour. Membranes were typically exposed for 1-2 days to either XAR-5 film (Kodak) or a PhosphorImager screen (Molecular Dynamics). The results are shown in FIG 10A-10E.

Hybridization with a probe from the TFIIA $\alpha/\beta$ -like region of SALF reveals a 1.8 kb mRNA that is present in testis, but not in other tissues (FIG 10A, lane 12). The isolation of the ALF cDNA which corresponds to this species is illustrated in FIG 7B. The predicted 3.8 kb SALF mRNA is not visible in mRNA from any of the tissues examined, including placenta, liver, and testis from which SALF can be amplified by PCR (FIG 7A and C). These results indicate that ALF, and TFIIA $\alpha/\beta$ , are the major transcripts encoding human TFIIA large subunits, and that SALF is relatively rare. Hybridization with a probe specific for the 5'-end of SALF (5'-SALF) reveals a 6.5 kb species that is present at highest levels in heart, placenta, kidney, prostate, and uterus (FIG 10B, lanes 1, 3, 7, 11, and 13) and at lower levels in other tissues. This transcript, termed RNA6.5, was not detected using the ALF-specific probe (FIG 10A), indicating that it does not contain a downstream ALF domain. Thus, RNA6.5 is an independent human transcript that contains sequences similar, or identical, to those present at the 5'-end of SALF.

A dot blot purchased from Clontech containing 89-514 ng poly(A) mRNA from 50 adult and fetal tissues is hybridized as follows: [ $\alpha$ 32P]-dCTP-labeled DNA is combined with 30  $\mu$ g Cot1 DNA (Boehringer-Mannheim) and 100  $\mu$ g salmon sperm DNA, denatured, and allowed to renature in 200 ml of 5X SSC at 68° C for 30 minutes prior to addition. After hybridization in 5 ml ExpressHyb solution at 65° C overnight with the probe, the blot was washed in 0.1X SSC at 55° C. Membranes were exposed as follows: FIG 11A, 19 hours; 11B, 2 hours 45 minutes; 11C, 14 hours; 11D, 25 hours, and 11E, 30 minutes. For

reprobing, Northern and dot blots were stripped twice with 0.5% SDS at 100° C, cooled to room temperature, and exposed overnight to confirm the loss of the previous signals. Quantitation of hybridization signals is performed using ImageQuaNT (Molecular Dynamics), and relative transcript levels in testis are determined by comparison to an average level from non-testis tissues. Ubiquitin (FIG 11E) is a control provided by the manufacturer (Clontech).

Using an ALF-specific probe, a strong signal was observed in testis that is due to the presence of the 1.8 kb ALF transcript (FIG 11A, position D1). In addition, weak signals were observed in approximately 24 of the remaining tissues, including small intestine, bladder, uterus, and prostate (positions E3, C5, C6, and C7). These signals indicate that ALF, or SALF, is expressed to low levels in non-testis tissues, and their detection in this study reflects the greater sensitivity of the dot blot. When this blot was stripped and reprobed with the 5'-SALF probe, signals were detected in all tissues (FIG 11B), with highest levels in placenta, uterus, spinal cord, and fetal kidney (positions F4, C6, B7, and G3) and several others, and lower levels in the remaining tissues. Because this probe detected high levels of RNA6.5 (but not SALF) in Northern analysis, the signals in Figure 6B are primarily due to the expression of RNA6.5.

Further inspection of the data reveals that the signals detected with the ALF- specific probe in FIG 11A are present in a range of tissues that is nearly identical to those observed in FIG 11B. This holds true for approximately 20 tissues, including bladder, uterus, prostate, ovary, placenta (positions C5, C6, C7, D2, and F4), and others, but not for testis (position D1). Likewise, the absence of signals in Figure 6A correlates with the absence of signals in FIG 11B. The results suggest a relationship between the expression of RNA6.5 with ALF-containing transcripts (possibly SALF) present at low levels in non-testis tissues. Hybridization with human TFIIA $\alpha/\beta$ - and TFIIA $\gamma$ -specific probes (FIG 11C and 11D) shows that the corresponding mRNAs are expressed in all tissues. Quantitation of the results confirms that ALF (50-fold), TFIIA $\alpha/\beta$  (4-fold), and TFIIA $\gamma$  (10-fold) are enriched in testis tissue.

EXAMPLE 4

**PRODUCTION OF ALF AND SALF POLYPEPTIDES**

To prepare recombinant ALF protein for functional assays, a 479 amino acid histidine-tagged polypeptide that spans residues Val7 to Trp478 was overexpressed and purified. Primers used in these studies had the corresponding sequences as follows: A1 (5'-ACTACTCATATGGCACACCACCATCACCACCATGTACCTAAACTCTACAGATC T-3') (SEQ ID NO.: 14) and A2 (5'-AGTAGTGGATCCTTACCACTCTGCATCACC-3") (SEQ ID NO.: 15) were used to create a 1,445 bp NdeI-BamHI PCR fragment whose reading frame begins with the N-terminal extension MHHHHHHV (SEQ ID NO.: 16) and terminates with the natural TAA stop codon. This construct does not encode the first six amino acids (MACLNP, SEQ ID NO.: 17) found in the intact testis-derived ALF cDNA. After subcloning into pRSETC (Invitrogen), the resulting construct (pRSET-ALF) was transformed into *E. coli* BL21(DE3)pLysS (Novagen) and was expressed and purified essentially as follows. Cells were grown in LB media at 37° C to an OD<sub>600</sub> of ~0.5, and production of the 69 kD recombinant ALF protein is induced with 2 mM IPTG. Cells were harvested 3 hours post-induction, solubilized in Buffer A (0.1 M NaH2PO4, 0.01 M Tris pH 8.0, and either 6 M guanidine or 8 M urea), and sonicated five times for 30 seconds. The denatured cell lysate (~20 ml) was incubated with 2 ml Ni-NTA agarose resin (Qiagen) at room temperature for 1 hour. The resin was washed successively with Buffer A containing 8 M urea at pH 8.0, 6.3, and 5.9, and bound polypeptides are eluted at pH 3.5. Preparation of expression constructs for rat TFIIA $\alpha/\beta$  and rat TFIIA $\gamma$  subunits (Genbank Accession numbers AF000943 and AF000944, respectively) and purification of the corresponding 55 and 12 kD recombinant proteins were performed. For transcription studies the recombinant p69 and p12 proteins were codialyzed in order to prevent precipitation of the p12 subunit.

The predicted size of this polypeptide is 53 kD, but the mobility on SDS-PAGE is 69 kD (FIG 12A, lane 2). This observation is similar to results showing that the predicted 42 kD product of hTFIIA $\alpha/\beta$  migrates at 55 kD (3,4), and may be due to the effect of charged region III. The mobilities of the purified recombinant rat TFIIA $\alpha/\beta$  (p55) and TFIIA $\gamma$  (p12)

subunits used in these studies are also shown in FIG 12A (lanes 3 and 4). These polypeptides are at least 98% identical to their human counterparts.

A related procedure is used to prepare human TBP. Specifically, TBP was expressed in *E. coli* BL21(DE3)pLysS (Novagen) from a pET11d (Novagen) vector that contains a histidine tagged TBP open reading frame, induced with 2 mM IPTG at OD600 0.3 and purified at 4° C from the soluble fraction of the bacterial lysate over Ni-NTA agarose. Purification was performed by washing the resin with D700 buffer (20 mM HEPES, 20% glycerol, 0.2 mM EDTA, 10 mM β-mercaptoethanol, 0.5 mM PMSF, and 700 mM KCl) that contained 5 mM, 10 mM and 15 mM imidazole, and eluting bound polypeptides with D700 buffer that contained 100 mM imidazole. Recombinant proteins were dialyzed against Buffer C (10 mM Tris pH 7.9, 2 mM DTT, 20% glycerol, and 0.5 mM PMSF) containing 100 mM KCl prior to use.

To express SALF in a rabbit reticulocyte lysate system, primers NN1 (5'-TACTGCTCGAGCAACTTAGAGT-3') (SEQ ID NO.: 18) and 2a2-8 were used to generate a 2,988 bp product from pRACE4. An internal 2,207 bp XhoI-BglII fragment (aa 1-716) derived from this PCR product was then inserted into the XhoI-BglII digested pT7T3D vector that contains EST ID259637. Because an internal BglII-BglII fragment that spans aa 717-1,084 was excised during preparation of this vector, this fragment was later reinserted in the appropriate orientation to create a full-length SALF ORF (pT7T3-SALF). This construct (0.8 mg) was used to program rabbit reticulocyte lysates in the presence of [<sup>35</sup>S]-methionine as described by the manufacturer (Promega). Labeled polypeptides are separated on 8% SDS-PAGE gels, and visualized by autoradiography. To determine whether this full-length SALF cDNA construct is capable of directing the translation of an intact protein, *in vitro* transcription-translation reactions were performed.

As shown in FIG 12D, rabbit reticulocyte lysates programmed with pT7T3-SALF produced an [<sup>35</sup>S]-methionine- labeled polypeptide that migrated at 170 kD (compared to a predicted size of 132 kD). Lysates programmed with pT7T3-SALF truncated at an internal EcoRI site at nucleotide position 960 produced a 36 kD product similar to the predicted size of 32 kD (data not shown). These results demonstrate that although SALF contains a

suboptimal ATG initiation codon (AAGATGT) and encodes a large ORF composed of two distinct regions, it can be translated efficiently.

## EXAMPLE 5

### ALF IS A TRANSCRIPTION FACTOR FOR RNA POLYMERASE II.

Functionality of ALF was demonstrated in TBP-dependent mobility shift assays and in RNA polymerase II-dependent *in vitro* transcription assays. Mobility shift assays were performed using 10 fmol of a [ $\gamma$ -<sup>32</sup>P] ATP kinase-labeled TATA-containing an oligonucleotide that spans nucleotides -40 to -16 of the Adenovirus Major Late (AdML) promoter. Binding reactions (25  $\mu$ l final volume) were performed in 10 mM HEPES (pH 7.9), 2% (wt/vol) PEG-8000, 60 mM KCl, 5 mM DTT, 0.2 mM EDTA, 5 mM ammonium sulfate, 4 mM MgCl<sub>2</sub>, and 8% glycerol. Recombinant rat p55 (30 ng; 29 nM), rat p12 (1.1 ug; 3.5 uM), human ALF (180 ng; 137 nM) and human TBP (125 ng; 133 nM) were added to reactions as indicated. Reactions were incubated for 30 minutes at room temperature, and complexes are separated on native 5% polyacrylamide gels containing 0.5X TBE and 5% glycerol. Competition experiments contain either cold AdML TATA or SP1 oligonucleotides, and antibody supershift reactions contain 2-4  $\mu$ l of rabbit polyclonal antiserum raised against the 55 kD hTFIIA $\alpha/\beta$  polypeptide (3).

The activity of polypeptides was tested in electrophoretic mobility shift assays under conditions in which TATA-Binding Protein (TBP) alone is unable to bind DNA (FIG 12B, lane 1). The presence of TFIIA $\alpha/\beta$  (p55) and TFIIA $\gamma$  (p12) stabilized the TBP-DNA interaction via TFIIA $\alpha/\beta/\gamma$ -TBP-DNA complex formation (lane 2). Likewise, the recombinant ALF (p69) polypeptide, in conjunction with the TFIIA $\gamma$  (p12) subunit, was able to form ALF/ $\gamma$ - TBP-DNA complexes (lane 5). Formation of this complex depends on the presence of both ALF and TFIIA $\gamma$ . Although ALF (p69) is 102 amino acids longer than TFIIA $\alpha/\beta$  (p55) and migrates as a larger species in SDS-PAGE, the ALF/ $\gamma$ -TBP- DNA complex migrates slightly faster than the TFIIA $\alpha/\beta/\gamma$ -TBP-DNA complex (lanes 2 and 5). These reactions are run side-by-side on the same gel using ALF and TFIIA subunits that have been purified and renatured using the same procedure. The specificity of ALF/ $\gamma$ -TBP-DNA

complexes is similar to TFIIA $\alpha$ / $\beta$ / $\gamma$ -TBP-DNA complexes, as judged by competition with specific TATA (lanes 3 and 6) and nonspecific Sp1-site (lanes 4 and 7) oligonucleotides. In addition, both complexes are supershifted to the well when co-incubated with antiserum against hTFIIA $\alpha$ / $\beta$  (lanes 8 and 9), indicating that ALF and TFIIA $\alpha$ / $\beta$  are immunologically related, and are present in the respective complexes.

To demonstrate that ALF is a functional polypeptide that regulates gene expression, TFIIA-dependent *in vitro* transcription assays are performed. For this purpose, advantage was taken of the fact that TFIIA $\alpha$ / $\beta$  contains an intrinsic seven-histidine region that allows for the efficient removal of TFIIA from HeLa cell nuclear extracts using Ni-NTA agarose. In brief, 200  $\mu$ l of extract were incubated with 100  $\mu$ l Ni-NTA agarose resin for 30 minutes at 4 °C in the presence of 400 mM KCl. Control extracts were processed similarly, except that no Ni-NTA agarose was present. After microcentrifugation for 5 minutes, the supernatants were removed and dialyzed for 3 hours against Buffer C that contains 100 mM KCl. Transcription reactions were performed using a template (pMLC2AT) that contains the AdML promoter upstream of a G-free cassette. The template was linearized at a SmaI site just beyond the G-free cassette prior to use. Each reaction (20  $\mu$ l) contained: 8  $\mu$ l nuclear extract (~60  $\mu$ g protein), 2  $\mu$ l (550 ng) of recombinant p69 (0.22  $\mu$ M) and p12 (0.9  $\mu$ M) proteins, 1  $\mu$ g pMLC2AT, 10 mM HEPES (pH 7.5), 25 mM KCl, 6 mM MgCl<sub>2</sub>, 625  $\mu$ M UTP, 625  $\mu$ M ATP, 35  $\mu$ M CTP, 200  $\mu$ M O-methyl-GTP, 3% glycerol, 0.7  $\mu$ l [ $\alpha$ -32P] CTP and 37.3 units of RNAGuard (Pharmacia). After incubation at 30° C for 45 minutes, the reactions were terminated by adding 270  $\mu$ l stop solution (0.25 M NaCl, 1% SDS, 20 mM Tris pH 7.5, 5 mM EDTA and 66.7  $\mu$ g/ml tRNA) and extracted with an equal volume of 1:1 phenol/chloroform. Ethanol precipitated transcripts were resuspended in formamide-containing loading dye and electrophoresed on 5% acrylamide gels containing 1X TBE and 8 M urea. Depleted extracts were transcriptionally inactive, but were restored to normal activity by the addition of TFIIA. As shown in FIG 12C (lane 1), control (undepleted) extracts produce a [ $\gamma$ <sup>32</sup>P]-CTP labeled G-free RNA transcript were expressed under the control of the AdML promoter (pMLC2AT). The addition of recombinant ALF (p69) and TFIIA $\gamma$  (p12) to these extracts did not enhance transcription (lane 2).

TFIIA-depleted extracts were transcriptionally inactive, and were not affected by the re-addition of either ALF (p69) or TFIIA $\gamma$  (p12) alone (lanes 3-5). The addition of both ALF (p69) and TFIIA $\gamma$  (p12), however, restored transcription to the level observed with control extracts (lane 6). The results of the electrophoretic mobility shift and in vitro transcription assays shown in FIG 12B and 12C, demonstrate that ALF has TFIIA $\alpha/\beta$ -like functional activity via TBP, and that both ALF and TFIIA $\alpha/\beta$  require TFIIA $\gamma$ , or a functionally similar subunit, for activity.

While this invention has been described in reference to illustrative embodiments, this description is not intended to be construed in a limiting sense. Various modifications and combinations of the illustrative embodiments, as well as other embodiments of the invention, will be apparent to persons skilled in the art upon reference to the description. It is therefore intended that the appended claims encompass any such modifications or embodiments.

SEQUENCE LISTING

5       <110> DeJong Dr., Jeff

10      <120> Transcription Factors Related to TFIIA

15      <130> 119941-1053

20      <140>

25      <141>

30      <160> 20

35      <170> PatentIn Ver. 2.0

40      <210> 1

45      <211> 1617

50      <212> DNA

55      <213> Homo sapiens

60      <220>

65      <221> CDS

<222> (16)..(1449)

5 <400> 1

gctggagggtg ctgtc atg gcc tgc ctc aac ccg gtg cct aaa ctc tac aga 51

Met Ala Cys Leu Asn Pro Val Pro Lys Leu Tyr Arg

10

15

tct gta att gaa gat gta att gaa gga gtt cgg aat cta ttt gct gaa 99

Ser Val Ile Glu Asp Val Ile Glu Gly Val Arg Asn Leu Phe Ala Glu

15                    20                    25

20

gaa ggt ata gag gaa caa gtt tta aaa gac ttg aag cag ctc tgg gaa 147

25 Glu Gly Ile Glu Glu Gln Val Leu Lys Asp Leu Lys Gln Leu Trp Glu

30                          35                          40

30

acc aag gtt ttg cag tct aaa gca aca gaa gac ttc ttc aqa aat aqg 195

Thr Lys Val Leu Gln Ser Lys Ala Thr Glu Asp Phe Phe Arg Asn Ser

35

45                    50                    55                    60

atc caa tca cct ctg ttt act ctt cag ttg ccg cac agc ttg cac caa 243

40

Ile Gln Ser Pro Leu Phe Thr Leu Gln Leu Pro His Ser Leu His Gln

65                    70                    75

aca ttg caa tcg tca aca gca tca tta gtt att cct gct ggt aga act 291

5 Thr Leu Gln Ser Ser Thr Ala Ser Leu Val Ile Pro Ala Gly Arg Thr

80 . 85 90

10

ctt cca agt ttt acc aca gca gaa ctg ggc act tca aac tcc agt gca 339

Leu Pro Ser Phe Thr Thr Ala Glu Leu Gly Thr Ser Asn Ser Ser Ala

15

95 100 105

20

aac ttt act ttt cct ggt tat ccc att cat gta cca gca ggt gtg aca 387

Asn Phe Thr Phe Pro Gly Tyr Pro Ile His Val Pro Ala Gly Val Thr

110 115 120

25

cta cag act gta tct ggt cac ctt tat aaa gtc aat gta cca att atg 435

Leu Gln Thr Val Ser Gly His Leu Tyr Lys Val Asn Val Pro Ile Met

30

125 130 135 140

35

gtg aca gag act tct gga aga gca ggt att ctt cag cat cca att cag 483

Val Thr Glu Thr Ser Gly Arg Ala Gly Ile Leu Gln His Pro Ile Gln

40

145 150 155

caa gta ttt caa cag ctt ggc cag cct tca gta ata caa act agt gtt 531

Gln Val Phe Gln Gln Leu Gly Gln Pro Ser Val Ile Gln Thr Ser Val

160 165 170

5

cca caa ttg aat cca tgg tct ctt .caa gca act act gaa aaa tca cag 579

Pro Gln Leu Asn Pro Trp Ser Leu Gln Ala Thr Thr Glu Lys Ser Gln  
10

175 180 185

15 aga att gaa acc gtg cta cag caa ccc gca att cta cct tct ggg cca 627

Arg Ile Glu Thr Val Leu Gln Gln Pro Ala Ile Leu Pro Ser Gly Pro

190 195 200

20

gta gat agg aaa cac tta gaa aat gcc acc agt gat ata ctt gta tct 675

25

Val Asp Arg Lys His Leu Glu Asn Ala Thr Ser Asp Ile Leu Val Ser

205 210 215 220

30

cct gga aat gag cat aaa atc gtg cct gaa gct ttg ttg tgt cat cag 723

Pro Gly Asn Glu His Lys Ile Val Pro Glu Ala Leu Leu Cys His Gln

35

225 230 235

gaa agt tct cac tat atc agt ctt cca ggt gtt gta ttt tct cca cag 771

40

Glu Ser Ser His Tyr Ile Ser Leu Pro Gly Val Val Phe Ser Pro Gln

240 245 250

gtc tct caa aca aat tct gat gtg gag tca gtg ctc agt ggt tca gct 819

5 Val Ser Gln Thr Asn Ser Asp Val Glu Ser Val Leu Ser Gly Ser Ala

255 260 265

10

agc atg gct caa aat ctg cat gat gag tcc ctc tcc aca agc cct cat 867

Ser Met Ala Gln Asn Leu His Asp Glu Ser Leu Ser Thr Ser Pro His

15 270 275 280

20 ggg gct ctc cac cag cac gtg act gat att cag ctt cat att ctt aaa 915

Gly Ala Leu His Gln His Val Thr Asp Ile Gln Leu His Ile Leu Lys

285 290 295 300

25

aat agg atg tat gga tgt gat tct gta aag caa cca aga aat ata gag 963

Asn Arg Met Tyr Gly Cys Asp Ser Val Lys Gln Pro Arg Asn Ile Glu

30 305 310 315

35 gaa ccc agc aac ata cct gta tca gag aag gat tct aat tct cag gtg 1011

Glu Pro Ser Asn Ile Pro Val Ser Glu Lys Asp Ser Asn Ser Gln Val

320 325 330

40

gat tta agc att cggtt act gat gat gat att ggt gaa ata att caa 1059

Asp Leu Ser Ile Arg Val Thr Asp Asp Asp Ile Gly Glu Ile Ile Gln

335

340

345

5

gta gat gga agc ggt gat aca tct.tcc aat gaa gaa ata gga agt aca 1107

Val Asp Gly Ser Gly Asp Thr Ser Ser Asn Glu Glu Ile Gly Ser Thr

10

350

355

360

15

aga gat gca gat gag aat gaa ttt cta ggg aat att gac ggg gga gat 1155

Arg Asp Ala Asp Glu Asn Glu Phe Leu Gly Asn Ile Asp Gly Gly Asp

20

365

370

375

380

25

ctg aag gta cct gaa gaa gct gac agt att tca aat gag gat tca 1203

30

Leu Lys Val Pro Glu Glu Ala Asp Ser Ile Ser Asn Glu Asp Ser

35

385

390

395

40

gcc aca aac agt agt gat aat gaa gac cct caa gta aac att gta gaa 1251

Ala Thr Asn Ser Ser Asp Asn Glu Asp Pro Gln Val Asn Ile Val Glu

35

400

405

410

gag gac cct tta aat tct gga gat gat gtt agt gaa cag gat gtg cca 1299

40

Glu Asp Pro Leu Asn Ser Gly Asp Asp Val Ser Glu Gln Asp Val Pro

415

420

425

gac ctg ttt gac acg gat aat gta att gtc tgt cag tat gat aag att 1347

5 Asp Leu Phe Asp Thr Asp Asn Val Ile Val Cys Gln Tyr Asp Lys Ile

430 435 440

10

cat cga agc aag aac aaa tgg aaa ttc tat ttg aaa gat ggt gtt atg 1395

His Arg Ser Lys Asn Lys Trp Lys Phe Tyr Leu Lys Asp Gly Val Met

15 445 450 455 460

20tgt ttt gga ggg aga gac tat gta ttt gca aaa gcc att ggt gat gca 1443

Cys Phe Gly Gly Arg Asp Tyr Val Phe Ala Lys Ala Ile Gly Asp Ala

465 470 475

25

gag tgg taaaccttgc gagctcgat catctatttt gtgaacatca gttggactat 1499

30Glu Trp

35

attgcattt gtgaattcat ttttattttg aatatagtcc agcacagagc tgttcaaatt 1559

40tttagttcac tgcgtatggaat ttaataaaaat tataattcag atgcagatac aattacac 1617

<210> 2

<211> 478

<212> PRT

5 <213> Homo sapiens

10 <400> 2

Met Ala Cys Leu Asn Pro Val Pro Lys Leu Tyr Arg Ser Val Ile Glu

1 5 10 15

15

Asp Val Ile Glu Gly Val Arg Asn Leu Phe Ala Glu Glu Gly Ile Glu

20 25 30

Glu Gln Val Leu Lys Asp Leu Lys Gln Leu Trp Glu Thr Lys Val Leu

25 30 35 40 45

Gln Ser Lys Ala Thr Glu Asp Phe Phe Arg Asn Ser Ile Gln Ser Pro

30 35 40 45 50 55 60

35 Leu Phe Thr Leu Gln Leu Pro His Ser Leu His Gln Thr Leu Gln Ser

65 70 75 80

40

Ser Thr Ala Ser Leu Val Ile Pro Ala Gly Arg Thr Leu Pro Ser Phe

85 90 95

60

Thr Thr Ala Glu Leu Gly Thr Ser Asn Ser Ser Ala Asn Phe Thr Phe

5

100

105

110

10

Pro Gly Tyr Pro Ile His Val Pro Ala Gly Val Thr Leu Gln Thr Val

15

Ser Gly His Leu Tyr Lys Val Asn Val Pro Ile Met Val Thr Glu Thr

130

135

140

20

Ser Gly Arg Ala Gly Ile Leu Gln His Pro Ile Gln Gln Val Phe Gln

145

150

155

160

25

Gln Leu Gly Gln Pro Ser Val Ile Gln Thr Ser Val Pro Gln Leu Asn

30

165

170

175

35

Pro Trp Ser Leu Gln Ala Thr Thr Glu Lys Ser Gln Arg Ile Glu Thr

180

185

190

40

Val Leu Gln Gln Pro Ala Ile Leu Pro Ser Gly Pro Val Asp Arg Lys

195

200

205

His Leu Glu Asn Ala Thr Ser Asp Ile Leu Val Ser Pro Gly Asn Glu

210 215 220

5

His Lys Ile Val Pro Glu Ala Leu Leu Cys His Gln Glu Ser Ser His

225 230 235 240

10

Tyr Ile Ser Leu Pro Gly Val Val Phe Ser Pro Gln Val Ser Gln Thr

15 245 250 255

Asn Ser Asp Val Glu Ser Val Leu Ser Gly Ser Ala Ser Met Ala Gln  
20

260 265 270

25 Asn Leu His Asp Glu Ser Leu Ser Thr Ser Pro His Gly Ala Leu His

275 280 285

30 Gln His Val Thr Asp Ile Gln Leu His Ile Leu Lys Asn Arg Met Tyr  
290 295 300

35

Gly Cys Asp Ser Val Lys Gln Pro Arg Asn Ile Glu Glu Pro Ser Asn

305 310 315 320

40

Ile Pro Val Ser Glu Lys Asp Ser Asn Ser Gln Val Asp Leu Ser Ile

325 330 335

5 Arg Val Thr Asp Asp Asp Ile Gly Glu Ile Ile Gln Val Asp Gly Ser

340 345 350

10 Gly Asp Thr Ser Ser Asn Glu Glu Ile Gly Ser Thr Arg Asp Ala Asp

355 360 365

15 Glu Asn Glu Phe Leu Gly Asn Ile Asp Gly Asp Leu Lys Val Pro

20 370 375 380

25 Glu Glu Glu Ala Asp Ser Ile Ser Asn Glu Asp Ser Ala Thr Asn Ser

30 385 390 395 400

35 Ser Asp Asn Glu Asp Pro Gln Val Asn Ile Val Glu Asp Pro Leu

405 410 415

35 Asn Ser Gly Asp Asp Val Ser Glu Gln Asp Val Pro Asp Leu Phe Asp

420 425 430

40 Thr Asp Asn Val Ile Val Cys Gln Tyr Asp Lys Ile His Arg Ser Lys

435 440 445

Asn Lys Trp Lys Phe Tyr Leu Lys Asp Gly Val Met Cys Phe Gly Gly

5 450 455 460

Arg Asp Tyr Val Phe Ala Lys Ala Ile Gly Asp Ala Glu Trp  
10 465 470 475

15

<210> 3

20 <211> 3824

<212> DNA

<213> Homo sapiens

25 <220>

30 <221> CDS

<222> (115)..(3660)

35 <400> 3

ggactttggg actggacaga cctggtcaca gtctaggta tacatcttac tggtcgagca 60

40 actttagagt caacctattt gatttcttga caagaccaca atctgatccc aaag atg 117

Met

5 tgc tcc aca aat cca ggc aaa tgg gtc acc ttt gat gat gat cct gct 165  
Cys Ser Thr Asn Pro Gly Lys Trp Val Thr Phe Asp Asp Asp Pro Ala  
10 5 10 15  
gtt caa tct tct caa aag tca aag aat ttt cct ctg gag aat caa ggt 213  
15 Val Gln Ser Ser Gln Lys Ser Lys Asn Phe Pro Leu Glu Asn Gln Gly  
20 25 30  
gtc tgt aga cca aat gga ctg aag ctg aac cct cct ggc ctc agg gaa 261  
Val Cys Arg Pro Asn Gly Leu Lys Leu Asn Pro Pro Gly Leu Arg Glu  
25 30 35 40 45  
ttt ccc agt gga tct tcc tcc acc agc agc act cct ctc tcc tcc ccc 309  
30 Phe Pro Ser Gly Ser Ser Thr Ser Ser Thr Pro Leu Ser Ser Pro  
35 50 55 60 65  
att gta gat ttt tat ttc agt cca gga cct cca agt aac tct cct ctt 357  
Ile Val Asp Phe Tyr Phe Ser Pro Gly Pro Pro Ser Asn Ser Pro Leu  
40 70 75 80

tct aca cct acc aaa gac ttc cca ggt ttt cct ggc atc ccc aaa gca 405

Ser Thr Pro Thr Lys Asp Phe Pro Gly Phe Pro Gly Ile Pro Lys Ala

5

85

90

95

10

ggg act cat gtg ctt tat cct att cca gaa tca tct tca gac agc cca 453

15

ctc gca ata tca gga gga gaa tct tcc tta ctg cct acc aga cca aca 501

20

Leu Ala Ile Ser Gly Gly Glu Ser Ser Leu Leu Pro Thr Arg Pro Thr

115

120

125

25

tgt tta tcc cat gcc ttg tta ccc agt gac cac tca tgt aca cat cca 549

30

Cys Leu Ser His Ala Leu Leu Pro Ser Asp His Ser Cys Thr His Pro

130

135

140

145

35

act ccc aaa gta ggt ctt cca gat gaa gtt aat cct caa cag gct gaa 597

Thr Pro Lys Val Gly Leu Pro Asp Glu Val Asn Pro Gln Gln Ala Glu

150

155

160

40

agc cta gga ttc caa agt gat gat ctc ccc cag ttt cag tat ttt cga 645

Ser Leu Gly Phe Gln Ser Asp Asp Leu Pro Gln Phe Gln Tyr Phe Arg

165 170 175

5 gag gac tgt gct ttt tca agt cca ttt cgg aaa gat gaa ggc agt gat 693

Glu Asp Cys Ala Phe Ser Ser Pro Phe Arg Lys Asp Glu Gly Ser Asp

10 180 185 190

tcc cat ttc acc ctt gac cca cca gga agc aaa aag atg ttc tca tca 741

15 Ser His Phe Thr Leu Asp Pro Pro Gly Ser Lys Lys Met Phe Ser Ser

195 200 205

20 aga aac aag gag atg cct att gac caa aaa agc cta aat aag tgt tca 789  
Arg Asn Lys Glu Met Pro Ile Asp Gln Lys Ser Leu Asn Lys Cys Ser

25 210 215 220 225

30 ctc aac tat atc tgt gag aag ctt gaa cat ctc cag tca gct gag aac 837

Leu Asn Tyr Ile Cys Glu Lys Leu Glu His Leu Gln Ser Ala Glu Asn

35

caa gac tca ctt aga agt ttg tct atg cac tgt cta tgt gct gaa gaa 885

Gln Asp Ser Leu Arg Ser Leu Ser Met His Cys Leu Cys Ala Glu Glu

40

245 250 255

aat gcc tct tcc ttt gtc ccc cac aca ctc ttc agg agt cag cca aaa 933

Asn Ala Ser Ser Phe Val Pro His Thr Leu Phe Arg Ser Gln Pro Lys

5 260 265

270

10 tcc gga tgg tct ttc atg ctg aga att cct gag aag aag aat atg atg 981

Ser Gly Trp Ser Phe Met Leu Arg Ile Pro Glu Lys Lys Asn Met Met

275 280 285

15

tct tcc cgg caa tgg gga cca att ttt ctg aaa gtt ttg cct gga gga 1029

Ser Ser Arg Gln Trp Gly Pro Ile Phe Leu Lys Val Leu Pro Gly Gly

20

290 295 300 305

25

att ttg cag atg tat tat gaa cag gga tta gaa aaa cca ttt aaa gag 1077

Ile Leu Gln Met Tyr Tyr Glu Gln Gly Leu Glu Lys Pro Phe Lys Glu

30

310 315 320

ata cag ctt gat cca tat tgt agg ctt tct gaa ccc aag gtt gag aac 1125

35

Ile Gln Leu Asp Pro Tyr Cys Arg Leu Ser Glu Pro Lys Val Glu Asn

325 330 335

40

ttc agt gta gca gga aaa atc cac act gtg aag att gaa cat gtg tct 1173

Phe Ser Val Ala Gly Lys Ile His Thr Val Lys Ile Glu His Val Ser

5                   340                   345                   350

5        tac aca gaa aaa agg aaa tac cat tct aag aca gaa gta gtt cat gaa   1221  
Tyr Thr Glu Lys Arg Lys Tyr His Ser Lys Thr Glu Val Val His Glu

10                   355                   360                   365

10        cct gac ata gag cag atg ctg aag ttg ggg tcc aca tcg tac cat gac   1269  
Pro Asp Ile Glu Gln Met Leu Lys Leu Gly Ser Thr Ser Tyr His Asp

15                   370                   375                   380                   385

15        ttc ctt gac ttt ctg act act gtg gag gag gag ctg atg aag ttg cca   1317  
Phe Leu Asp Phe Leu Thr Thr Val Glu Glu Leu Met Lys Leu Pro

20                   390                   395                   400

20        gct gtt tca aaa cca aaa aag aac tac gag gag caa gaa att tcc ttg   1365  
Ala Val Ser Lys Pro Lys Lys Asn Tyr Glu Glu Gln Glu Ile Ser Leu

25                   405                   410                   415

25        gaa att gtg gac aac ttt tgg ggt aaa gtc aca aaa gaa gga aaa ttt   1413  
Glu Ile Val Asp Asn Phe Trp Gly Lys Val Thr Lys Glu Gly Lys Phe

30                   420                   425                   430

gtt gaa agt gct gtg ata act caa att tat tgc ctc tgc ttt gtg aat 1461

Val Glu Ser Ala Val Ile Thr Gln Ile Tyr Cys Leu Cys Phe Val Asn

5 435 440 445

10 ggg aac ctg gaa tgc ttt tta acc ttg aat gac ctt gag ttg ccg aag 1509

Gly Asn Leu Glu Cys Phe Leu Thr Leu Asn Asp Leu Glu Leu Pro Lys

15 450 455 460 465

cga gat gaa tcc tat tat gag aag gac tca gaa aaa aag ggg att gat 1557

20 Arg Asp Glu Ser Tyr Tyr Glu Lys Asp Ser Glu Lys Lys Gly Ile Asp

25 470 475 480

att ctt gac tac cat ttt cat aag tgt gtg aat gta caa gaa ttt gag 1605

Ile Leu Asp Tyr His Phe His Lys Cys Val Asn Val Gln Glu Phe Glu

30 485 490 495

caa tca aga atc att aag ttt gta cct ctg gat gcc tgc cgg ttt gag 1653

35 Gln Ser Arg Ile Ile Lys Phe Val Pro Leu Asp Ala Cys Arg Phe Glu

500 505 510

40 ctg atg cgt ttc aag act ttg tat aat ggg gat aat ctt ccc ttt tcc 1701

Leu Met Arg Phe Lys Thr Leu Tyr Asn Gly Asp Asn Leu Pro Phe Ser

515 520 525

5 ttg aag tct gta gtg gtt gtc cag gga gca tac gtg gaa ctt cag gct 1749  
Leu Lys Ser Val Val Val Val Gln Gly Ala Tyr Val Glu Leu Gln Ala.

10 530 535 540 545

ttt gtc aac atg gcc tca ttg gcg cag agg tca tcc tat gct ggt tcc 1797  
15 Phe Val Asn Met Ala Ser Leu Ala Gln Arg Ser Ser Tyr Ala Gly Ser

550 555 560

20 tta agg tcc tgt gac aat ata agg ata cac ttt cct gtc cca tcg cag 1845  
Leu Arg Ser Cys Asp Asn Ile Arg Ile His Phe Pro Val Pro Ser Gln

25 565 570 575

30 tgg atc aag gcc ctt tgg acc atg aac ctc cag agg cag aag tct ctg 1893  
Trp Ile Lys Ala Leu Trp Thr Met Asn Leu Gln Arg Gln Lys Ser Leu

35 580 585 590

aaa gct aaa atg aac cgc cga gca tgt ctg ggg agt tta cag gaa ctt 1941  
40 Lys Ala Lys Met Asn Arg Arg Ala Cys Leu Gly Ser Leu Gln Glu Leu

595 600 605

gaa tct gaa cct gtc att caa gtc act gtg ggg tca gca aaa tat gag 1989

Glu Ser Glu Pro Val Ile Gln Val Thr Val Gly Ser Ala Lys Tyr Glu

5 610 615 620 625

10 agt gcc tac cag gca gtg gta tgg aag ata gat cgg ctt cca gac aaa 2037

Ser Ala Tyr Gln Ala Val Val Trp Lys Ile Asp Arg Leu Pro Asp Lys

630 635 640

15

aat tca agt cta gat cat ccc cat tgt ctg tca tac aaa tta gag ctt 2085

Asn Ser Ser Leu Asp His Pro His Cys Leu Ser Tyr Lys Leu Glu Leu

20 645 650 655

25 gga tca gac caa gaa att ccc tct gat tgg tat cca ttt gct act gtt 2133

Gly Ser Asp Gln Glu Ile Pro Ser Asp Trp Tyr Pro Phe Ala Thr Val

30 660 665 670

cag ttt tcc gtg cct gac acc tgt gcc tca agg aca gag gtc agg tct 2181

35 Gln Phe Ser Val Pro Asp Thr Cys Ala Ser Arg Thr Glu Val Arg Ser

675 680 685

40

ctg gga gtg gag agt gat gtc cag cca cag aaa cat gtt cag cag cga 2229

Leu Gly Val Glu Ser Asp Val Gln Pro Gln Lys His Val Gln Gln Arg

690 695 700 705

5 gct tgc tac aac atc cag cct aaa ctc tac aga tct gta att gaa gat 2277

Ala Cys Tyr Asn Ile Gln Pro Lys Leu Tyr Arg Ser Val Ile Glu Asp

10 710 715 720

gta att gaa gga gtt cg<sup>g</sup> aat cta ttt gct gaa gaa ggt ata gag gaa 2325

15 Val Ile Glu Gly Val Arg Asn Leu Phe Ala Glu Glu Gly Ile Glu Glu

725 730 735

20 caa gtt tta aaa gac ttg aag cag ctc tgg gaa acc aag gtt ttg cag 2373  
Gln Val Leu Lys Asp Leu Lys Gln Leu Trp Glu Thr Lys Val Leu Gln

25 740 745 750

30 tct aaa gca aca gaa gac ttc ttc aga aat agc atc caa tca cct ctg 2421

Ser Lys Ala Thr Glu Asp Phe Phe Arg Asn Ser Ile Gln Ser Pro Leu

755 760 765

35

ttt act ctt cag ttg ccg cac agc ttg cac caa aca ttg caa tcg tca 2469

40 Phe Thr Leu Gln Leu Pro His Ser Leu His Gln Thr Leu Gln Ser Ser

770 775 780 785

aca gca tca tta gtt att cct gct ggt aga act ctt cca agt ttt acc 2517

Thr Ala Ser Leu Val Ile Pro Ala Gly Arg Thr Leu Pro Ser Phe Thr

5 790 795 800

aca gca gaa ctg ggc act tca aac tcc agt gca aac ttt act ttt cct 2565

10 Thr Ala Glu Leu Gly Thr Ser Asn Ser Ser Ala Asn Phe Thr Phe Pro

805 810 815

15

ggt tat ccc att cat gta cca gca ggt gtg aca cta cag act gta tct 2613

Gly Tyr Pro Ile His Val Pro Ala Gly Val Thr Leu Gln Thr Val Ser

20 820 825 830

25 ggt cac ctt tat aaa gtc aat gta cca att atg gtg aca gag act tct 2661

Gly His Leu Tyr Lys Val Asn Val Pro Ile Met Val Thr Glu Thr Ser

30 835 840 845

gga aga gca ggt att ctt cag cat cca att cag caa gta ttt caa cag 2709

35 Gly Arg Ala Gly Ile Leu Gln His Pro Ile Gln Gln Val Phe Gln Gln

850 855 860 865

40

ctt ggc cag cct tca gta ata caa act agt gtt cca caa ttg aat cca 2757

Leu Gly Gln Pro Ser Val Ile Gln Thr Ser Val Pro Gln Leu Asn Pro

870 875 880

5 tgg tct ctt caa gca act act gaa aaa tca cag aga att gaa acc gtg 2805

Trp Ser Leu Gln Ala Thr Thr Glu Lys Ser Gln Arg Ile Glu Thr Val

885 890 895

10

cta cag caa ccc gca att cta cct tct ggg cca gta gat agg aaa cac 2853

15

Leu Gln Gln Pro Ala Ile Leu Pro Ser Gly Pro Val Asp Arg Lys His

900 905 910

20

tta gaa aat gcc acc agt gat ata ctt gta tct cct gga aat gag cat 2901

Leu Glu Asn Ala Thr Ser Asp Ile Leu Val Ser Pro Gly Asn Glu His

25

915 920 925

30

aaa atc gtg cct gaa gct ttg ttg tgt cat cag gaa agt tct cac tat 2949

Lys Ile Val Pro Glu Ala Leu Leu Cys His Gln Glu Ser Ser His Tyr

930 935 940 945

35

atc agt ctt cca ggt gtt gta ttt tct cca cag gtc tct caa aca aat 2997

Ile Ser Leu Pro Gly Val Val Phe Ser Pro Gln Val Ser Gln Thr Asn

40

950 955 960

tct gat gtg gag tca gtg ctc agt ggt tca gct agc atg gct caa aat 3045

Ser Asp Val Glu Ser Val Leu Ser Gly Ser Ala Ser Met Ala Gln Asn

5

965

970

975

10

ctg cat gat gag tcc ctc tcc aca agc cct cat ggg gct ctc cac cag 3093

Leu His Asp Glu Ser Leu Ser Thr Ser Pro His Gly Ala Leu His Gln

980

985

990

15

cac gtg act gat att cag ctt cat att ctt aaa aat agg atg tat gga 3141

20

His Val Thr Asp Ile Gln Leu His Ile Leu Lys Asn Arg Met Tyr Gly

995

1000

1005

25

tgt gat tct gta aag caa cca aga aat ata gag gaa ccc agc aac ata 3189

Cys Asp Ser Val Lys Gln Pro Arg Asn Ile Glu Glu Pro Ser Asn Ile

30

1010

1015

1020

1025

35

cct gta tca gag aag gat tct aat tct cag gtg gat tta agc att cg 3237

Pro Val Ser Glu Lys Asp Ser Asn Ser Gln Val Asp Leu Ser Ile Arg

1030

1035

1040

40

gtt act gat gat att ggt gaa ata att caa gta gat gga agc ggt 3285

Val Thr Asp Asp Asp Ile Gly Glu Ile Ile Gln Val Asp Gly Ser Gly

1045 1050 1055

5 gat aca tct tcc aat gaa gaa ata gga agt aca aga gat gca gat gag 3333

Asp Thr Ser Ser Asn Glu Glu Ile Gly Ser Thr Arg Asp Ala Asp Glu

1060 1065 1070

10

aat gaa ttt cta ggg aat att gac ggg gga gat ctg aag gta cct gaa 3381

15 Asn Glu Phe Leu Gly Asn Ile Asp Gly Gly Asp Leu Lys Val Pro Glu

1075 1080 1085

20

gaa gaa gct gac agt att tca aat gag gat tca gcc aca aac agt agt 3429

Glu Glu Ala Asp Ser Ile Ser Asn Glu Asp Ser Ala Thr Asn Ser Ser

25

1090 1095 1100 1105

30

gat aat gaa gac cct caa gta aac att gta gaa gag gac cct tta aat 3477

Asp Asn Glu Asp Pro Gln Val Asn Ile Val Glu Glu Asp Pro Leu Asn

1110 1115 1120

35

tct gga gat gat gtt agt gaa cag gat gtg cca gac ctg ttt gac acg 3525

Ser Gly Asp Asp Val Ser Glu Gln Asp Val Pro Asp Leu Phe Asp Thr

40

1125 1130 1135

gat aat gta att gtc tgt cag tat gat aag att cat cga agc aag aac 3573

Asp Asn Val Ile Val Cys Gln Tyr Asp Lys Ile His Arg Ser Lys Asn

5

1140

1145

1150

10

aaa tgg aaa ttc tat ttg aaa gat ggt gtt atg tgt ttt gga ggg aga 3621

Lys Trp Lys Phe Tyr Leu Lys Asp Gly Val Met Cys Phe Gly Arg

15

1155

1160

1165

20  
25  
30

gac tat gta ttt gca aaa gcc att ggt gat gca gag tgg taaaccttgt 3670

Asp Tyr Val Phe Ala Lys Ala Ile Gly Asp Ala Glu Trp

1170

1175

1180

25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000  
1001  
1002  
1003  
1004  
1005  
1006  
1007  
1008  
1009  
1000  
1001  
1002  
1003  
1004  
1005  
1006  
1007  
1008  
1009  
1010  
1011  
1012  
1013  
1014  
1015  
1016  
1017  
1018  
1019  
1010  
1011  
1012  
1013  
1014  
1015  
1016  
1017  
1018  
1019  
1020  
1021  
1022  
1023  
1024  
1025  
1026  
1027  
1028  
1029  
1020  
1021  
1022  
1023  
1024  
1025  
1026  
1027  
1028  
1029  
1030  
1031  
1032  
1033  
1034  
1035  
1036  
1037  
1038  
1039  
1030  
1031  
1032  
1033  
1034  
1035  
1036  
1037  
1038  
1039  
1040  
1041  
1042  
1043  
1044  
1045  
1046  
1047  
1048  
1049  
1040  
1041  
1042  
1043  
1044  
1045  
1046  
1047  
1048  
1049  
1050  
1051  
1052  
1053  
1054  
1055  
1056  
1057  
1058  
1059  
1050  
1051  
1052  
1053  
1054  
1055  
1056  
1057  
1058  
1059  
1060  
1061  
1062  
1063  
1064  
1065  
1066  
1067  
1068  
1069  
1060  
1061  
1062  
1063  
1064  
1065  
1066  
1067  
1068  
1069  
1070  
1071  
1072  
1073  
1074  
1075  
1076  
1077  
1078  
1079  
1070  
1071  
1072  
1073  
1074  
1075  
1076  
1077  
1078  
1079  
1080  
1081  
1082  
1083  
1084  
1085  
1086  
1087  
1088  
1089  
1080  
1081  
1082  
1083  
1084  
1085  
1086  
1087  
1088  
1089  
1090  
1091  
1092  
1093  
1094  
1095  
1096  
1097  
1098  
1099  
1090  
1091  
1092  
1093  
1094  
1095  
1096  
1097  
1098  
1099  
1100  
1101  
1102  
1103  
1104  
1105  
1106  
1107  
1108  
1109  
1100  
1101  
1102  
1103  
1104  
1105  
1106  
1107  
1108  
1109  
1110  
1111  
1112  
1113  
1114  
1115  
1116  
1117  
1118  
1119  
1110  
1111  
1112  
1113  
1114  
1115  
1116  
1117  
1118  
1119  
1120  
1121  
1122  
1123  
1124  
1125  
1126  
1127  
1128  
1129  
1120  
1121  
1122  
1123  
1124  
1125  
1126  
1127  
1128  
1129  
1130  
1131  
1132  
1133  
1134  
1135  
1136  
1137  
1138  
1139  
1130  
1131  
1132  
1133  
1134  
1135  
1136  
1137  
1138  
1139  
1140  
1141  
1142  
1143  
1144  
1145  
1146  
1147  
1148  
1149  
1140  
1141  
1142  
1143  
1144  
1145  
1146  
1147  
1148  
1149  
1150  
1151  
1152  
1153  
1154  
1155  
1156  
1157  
1158  
1159  
1150  
1151  
1152  
1153  
1154  
1155  
1156  
1157  
1158  
1159  
1160  
1161  
1162  
1163  
1164  
1165  
1166  
1167  
1168  
1169  
1160  
1161  
1162  
1163  
1164  
1165  
1166  
1167  
1168  
1169  
1170  
1171  
1172  
1173  
1174  
1175  
1176  
1177  
1178  
1179  
1170  
1171  
1172  
1173  
1174  
1175  
1176  
1177  
1178  
1179  
1180  
1181  
1182  
1183  
1184  
1185  
1186  
1187  
1188  
1189  
1180  
1181  
1182  
1183  
1184  
1185  
1186  
1187  
1188  
1189  
1190  
1191  
1192  
1193  
1194  
1195  
1196  
1197  
1198  
1199  
1190  
1191  
1192  
1193  
1194  
1195  
1196  
1197  
1198  
1199  
1200  
1201  
1202  
1203  
1204  
120

<213> Homo sapiens

5

<400> 4

Met Cys Ser Thr Asn Pro Gly Lys Trp Val Thr Phe Asp Asp Asp Pro

10

1 5 10 15

Ala Val Gln' Ser Ser Gln Lys Ser Lys Asn Phe Pro Leu Glu Asn Gln

15

15 20 25 30

Gly Val Cys Arg Pro Asn Gly Leu Lys Leu Asn Pro Pro Gly Leu Arg

20

20 25 30 35 40 45

Glu Phe Pro Ser Gly Ser Ser Ser Thr Ser Ser Thr Pro Leu Ser Ser

25

25 30 35 40 45 50 55 60

Pro Ile Val Asp Phe Tyr Phe Ser Pro Gly Pro Pro Ser Asn Ser Pro

30

30 35 40 45 50 55 60 65 70 75 80

100 105 110

5 Pro Leu Ala Ile Ser Gly Gly Glu Ser Ser Leu Leu Pro Thr Arg Pro

115 120 125

10

Thr Cys Leu Ser His Ala Leu Leu Pro Ser Asp His Ser Cys Thr His

130 135 140

15

Pro Thr Pro Lys Val Gly Leu Pro Asp Glu Val Asn Pro Gln Gln Ala

20 145 150 155 160

Glu Ser Leu Gly Phe Gln Ser Asp Asp Leu Pro Gln Phe Gln Tyr Phe

25 165 170 175

30 Arg Glu Asp Cys Ala Phe Ser Ser Pro Phe Arg Lys Asp Glu Gly Ser

35 180 185 190

Asp Ser His Phe Thr Leu Asp Pro Pro Gly Ser Lys Lys Met Phe Ser

40 195 200 205

Ser Arg Asn Lys Glu Met Pro Ile Asp Gln Lys Ser Leu Asn Lys Cys

210 215 220

Ser Leu Asn Tyr Ile Cys Glu Lys Leu Glu His Leu Gln Ser Ala Glu

5           225                   230                   235                   240

Asn Gln Asp Ser Leu Arg Ser Leu Ser Met His Cys Leu Cys Ala Glu

10                   245                   250                   255

15           Glu Asn Ala Ser Ser Phe Val Pro His Thr Leu Phe Arg Ser Gln Pro

              260                   265                   270

20                   Lys Ser Gly Trp Ser Phe Met Leu Arg Ile Pro Glu Lys Lys Asn Met

              275                   280                   285

25                   Met Ser Ser Arg Gln Trp Gly Pro Ile Phe Leu Lys Val Leu Pro Gly

              290                   295                   300

30                   Gly Ile Leu Gln Met Tyr Tyr Glu Gln Gly Leu Glu Lys Pro Phe Lys

35           305                   310                   315                   320

40                   Glu Ile Gln Leu Asp Pro Tyr Cys Arg Leu Ser Glu Pro Lys Val Glu

              325                   330                   335

Asn Phe Ser Val Ala Gly Lys Ile His Thr Val Lys Ile Glu His Val

340

345

350

5

Ser Tyr Thr Glu Lys Arg Lys Tyr His Ser Lys Thr Glu Val Val His

355

360

365

10

Glu Pro Asp Ile Glu Gln Met Leu Lys Leu Gly Ser Thr Ser Tyr His

15

370

375

380

20

Asp Phe Leu Asp Phe Leu Thr Thr Val Glu Glu Glu Leu Met Lys Leu

385

390

395

400

25

Pro Ala Val Ser Lys Pro Lys Lys Asn Tyr Glu Glu Gln Glu Ile Ser

30

405

410

415

Leu Glu Ile Val Asp Asn Phe Trp Gly Lys Val Thr Lys Glu Gly Lys

420

425

430

35

Phe Val Glu Ser Ala Val Ile Thr Gln Ile Tyr Cys Leu Cys Phe Val

435

440

445

40

Asn Gly Asn Leu Glu Cys Phe Leu Thr Leu Asn Asp Leu Glu Leu Pro

450 455 460

5 Lys Arg Asp Glu Ser Tyr Tyr Glu Lys Asp Ser Glu Lys Lys Gly Ile

465 470 475 480

10 Asp Ile Leu Asp Tyr His Phe His Lys Cys Val Asn Val Gln Glu Phe

485 490 495

15 Glu Gln Ser Arg Ile Ile Lys Phe Val Pro Leu Asp Ala Cys Arg Phe

20 500 505 510

25 Glu Leu Met Arg Phe Lys Thr Leu Tyr Asn Gly Asp Asn Leu Pro Phe

30 515 520 525

35 Ser Leu Lys Ser Val Val Val Gln Gly Ala Tyr Val Glu Leu Gln

30 530 535 540

35 Ala Phe Val Asn Met Ala Ser Leu Ala Gln Arg Ser Ser Tyr Ala Gly

40 545 550 555 560

45 Ser Leu Arg Ser Cys Asp Asn Ile Arg Ile His Phe Pro Val Pro Ser

565 570 575

Gln Trp Ile Lys Ala Leu Trp Thr Met Asn Leu Gln Arg Gln Lys Ser

5 580 585 590

Leu Lys Ala Lys Met Asn Arg Arg Ala Cys Leu Gly Ser Leu Gln Glu

10 595 600 605

15 Leu Glu Ser Glu Pro Val Ile Gln Val Thr Val Gly Ser Ala Lys Tyr

610 615 620

20 Glu Ser Ala Tyr Gln Ala Val Val Trp Lys Ile Asp Arg Leu Pro Asp

625 630 635 640

25 Lys Asn Ser Ser Leu Asp His Pro His Cys Leu Ser Tyr Lys Leu Glu

645 650 655

30 Leu Gly Ser Asp Gln Glu Ile Pro Ser Asp Trp Tyr Pro Phe Ala Thr

35 660 665 670

40 Val Gln Phe Ser Val Pro Asp Thr Cys Ala Ser Arg Thr Glu Val Arg

675 680 685

Ser Leu Gly Val Glu Ser Asp Val Gln Pro Gln Lys His Val Gln Gln

690 695 700

5

Arg Ala Cys Tyr Asn Ile Gln Pro Lys Leu Tyr Arg Ser Val Ile Glu

705 710 715 720

10

Asp Val Ile Glu Gly Val Arg Asn Leu Phe Ala Glu Glu Gly Ile Glu

15

725 730 735

20

Glu Gln Val Leu Lys Asp Leu Lys Gln Leu Trp Glu Thr Lys Val Leu

740 745 750

25

Gln Ser Lys Ala Thr Glu Asp Phe Phe Arg Asn Ser Ile Gln Ser Pro

755 760 765

30

Leu Phe Thr Leu Gln Leu Pro His Ser Leu His Gln Thr Leu Gln Ser

770 775 780

35

Ser Thr Ala Ser Leu Val Ile Pro Ala Gly Arg Thr Leu Pro Ser Phe

785 790 795 800

40

Thr Thr Ala Glu Leu Gly Thr Ser Asn Ser Ser Ala Asn Phe Thr Phe

805 810 815

5 Pro Gly Tyr Pro Ile His Val Pro Ala Gly Val Thr Leu Gln Thr Val

820 .825 830

10 Ser Gly His Leu Tyr Lys Val Asn Val Pro Ile Met Val Thr Glu Thr

835 840 845

15

Ser Gly Arg Ala Gly Ile Leu Gln His Pro Ile Gln Gln Val Phe Gln

20 850 855 860

Gln Leu Gly Gln Pro Ser Val Ile Gln Thr Ser Val Pro Gln Leu Asn

25 865 870 875 880

30 Pro Trp Ser Leu Gln Ala Thr Thr Glu Lys Ser Gln Arg Ile Glu Thr

885 890 895

35 Val Leu Gln Gln Pro Ala Ile Leu Pro Ser Gly Pro Val Asp Arg Lys

900 905 910

40

His Leu Glu Asn Ala Thr Ser Asp Ile Leu Val Ser Pro Gly Asn Glu

915 920 925

His Lys Ile Val Pro Glu Ala Leu Leu Cys His Gln Glu Ser Ser His

5 930 935 940

Tyr Ile Ser Leu Pro Gly Val Val Phe Ser Pro Gln Val Ser Gln Thr  
10 945 950 955 960

15 Asn Ser Asp Val Glu Ser Val Leu Ser Gly Ser Ala Ser Met Ala Gln  
965 970 975

20 Asn Leu His Asp Glu Ser Leu Ser Thr Ser Pro His Gly Ala Leu His  
980 985 990

25 Gln His Val Thr Asp Ile Gln Leu His Ile Leu Lys Asn Arg Met Tyr  
995 1000 1005

30 Gly Cys Asp Ser Val Lys Gln Pro Arg Asn Ile Glu Glu Pro Ser Asn  
35 1010 1015 1020

40 Ile Pro Val Ser Glu Lys Asp Ser Asn Ser Gln Val Asp Leu Ser Ile  
1025 1030 1035 1040

Arg Val Thr Asp Asp Asp Ile Gly Glu Ile Ile Gln Val Asp Gly Ser

1045

1050

1055

5

Gly Asp Thr Ser Ser Asn Glu Glu Ile Gly Ser Thr Arg Asp Ala Asp

1060

1065

1070

10

Glu Asn Glu Phe Leu Gly Asn Ile Asp Gly Gly Asp Leu Lys Val Pro

15

1075

1080

1085

20

Glu Glu Glu Ala Asp Ser Ile Ser Asn Glu Asp Ser Ala Thr Asn Ser

1090

1095

1100

25

Ser Asp Asn Glu Asp Pro Gln Val Asn Ile Val Glu Glu Asp Pro Leu

105

1110

1115

1120

30

Asn Ser Gly Asp Asp Val Ser Glu Gln Asp Val Pro Asp Leu Phe Asp

1125

1130

1135

35

Thr Asp Asn Val Ile Val Cys Gln Tyr Asp Lys Ile His Arg Ser Lys

1140

1145

1150

40

Asn Lys Trp Lys Phe Tyr Leu Lys Asp Gly Val Met Cys Phe Gly Gly

1155                    1160                    1165

5 Arg Asp Tyr Val Phe Ala Lys Ala Ile Gly Asp Ala Glu Trp

1170                    1175                    1180

10

<210> 5

15 <211> 261

<212> DNA

20 <213> Homo sapiens

25 <220>

30 <221> CDS

35 <222> (1)..(30)

40 <223> Variable 3' end for both ALF and SALF

<400> 5

gca ttc cca aga agg aca tcg ttt aac acc taaactcatt taacaaaagg'a        50

Ala Phe Pro Arg Arg Thr Ser Phe Asn Thr

1                    5                    10

40

tccgagaaga acagggacag tgtggaaaga aatcttcttg tgatggcata ttgccttcc 110

atatttcttc tggaatcatg ttcgcttggc ttccctgatta aaaacacagt tttattgctc 170

5

tctgcactgc caaaccaata aatttacaga agagaaagct gtattccact gtaccccttg 230

10

cagcatcaat aaaactgaca gccaaaaaaaaa a 261

15

<210> 6

20

<211> 10

25

<212> PRT

30

<213> Homo sapiens

35

<400> 6

40

Ala Phe Pro Arg Arg Thr Ser Phe Asn Thr

1

5

10

35

<210> 7

40

<211> 18

<212> DNA

<213> Homo sapiens

<400> 7

5 agaaattccc tctgattg

18

10

<210> 8

<211> 18

15

<212> DNA

<213> Homo sapiens

20

<400> 8

agtaacccga atgcttaa

18

25

<210> 9

<211> 18

<212> DNA

35

<213> Homo sapiens

<400> 9

40

atgctagctg aaccactg

18

<210> 10

5 <211> 27

<212> DNA

<213> Homo sapiens

10

<400> 10

15 ccatcctaat acgactcact atagggc

27

20

<210> 11

<211> 23

<212> DNA

<213> Homo sapiens

30

<400> 11

actcactata gggctcgagc ggc

23

35

40 <210> 12

<211> 30

<212> DNA

<213> Homo sapiens

5

<400> 12

ccagaaggta gaattgcggg ttgctgttagc

30

10

<210> 13

15

<211> 28

<212> DNA

20

<213> Homo sapiens

25

<400> 13

ggtgctgtca tggcctgcct caacccgg

28

30

<210> 14

35

<211> 52

<212> DNA

<213> Homo sapiens

40

<400> 14

actactcata tggcacacca tcaccatcac catgtaccta aactctacag at

52

5 <210> 15

<211> 30

<212> DNA

10 <213> Homo sapiens

15 <400> 15

agtagtggat ctttaccact ctgcattcacc

30

20  
25  
30

<210> 16

<211> 8

<212> PRT

<213> Homo sapiens

<400> 16

35 Met His His His His His Val

1

5

40

<210> 17

20  
25  
30  
35  
40

<211> 6

<212> PRT

5

<213> Homo sapiens

10  
15  
20  
25  
30  
35  
40

<400> 17

Met Ala Cys Leu Asn Pro

1

5

15

20  
25  
30  
35  
40

<210> 18

<211> 23

<212> DNA

<213> Homo sapiens

20  
25  
30  
35  
40

<400> 18

tactgctcga gcaacttttag agt

23

35

<210> 19

<211> 6

<212> PRT

<213> Homo sapiens

<400> 19

5 Ala Cys Leu Asn Pro Val

1 5

10

<210> 20

15 <211> 29

<212> DNA

20 <213> Homo sapiens

<400> 20

25 ggagtttgaa gtgcccaggt ctgctgtgg

29

30

## References

1. Roeder, R.G. (1991) *Trends Biochem. Sci.* 21, 327-335.
2. Orphanides, G., Lagrange, T., and Reinberg, D. (1996) *Genes Dev.* 10, 2657-2683.
3. DeJong, J. and Roeder, R.G. (1993) *Genes Dev.* 7, 2220-2234.
4. Ma, D., Watanabe, H., Mermelstein, F., Admon, A., Oguri, K., Sun, X., Wada, T., Imai, T., Shiroya, T., Reinberg, D., and Handa, H. (1993) *Genes Dev.* 7, 2246-2257.
5. Ozer, J., Moore, P.A., Bolden, A.H., Lee, A., Rosen, C.A., and Leiberman, P. (1994) *Genes Dev.* 8, 2324-2335.
10. Sun, X., Ma, D., Sheldon, M., Yeung, K., and Reinberg, D. (1994) *Genes Dev.* 8, 2336-2348.
11. DeJong, J., Bernstein, R., and Roeder, R.G. (1995) *Proc. Natl. Acad. Sci.* 92, 3313-3317.
15. Ranish, J.A., Lane, W.S., and Hahn, S. (1992) *Science* 255, 1127-1129.
16. Yokomori, K., Admon, A., Goodrich, J.A., Chen, J.-L., and Tjian, R. (1993) *Genes Dev.* 7, 2235-2245.
17. Yokomori, K., Zeidler, M.P., Chen, J.-L., Verrijzer, C.P., and Tjian, R. (1994) *Genes Dev.* 8, 2313-2323.
20. Bernstein, R., DeJong, J., and Roeder, R.G. (1994) *J. Biol. Chem.* 269, 24361-14366.
21. Tan, S., Hunziker, Y., Sargent, D.F., and Richmond, T.J. (1996) *Nature* 381, 127-134.
22. Geiger, J.H., Hahn, S., Lee, S., and Sigler, P.B. (1996) *Science* 272, 830-836.
23. Buratowski, S., Hahn, S., Sharp, P.A., and Guarente, L. (1989) *Cell* 56, 549-561.
24. Lee, D.K., DeJong, J., Hashimoto, S., Horikoshi, M., and Roeder, R.G. (1992) *Mol. Cell. Biol.* 12, 5189-5196.
25. Imbalzano, A.N., Zaret, K.S., and Kingston, R.E. (1994) *J. Biol. Chem.* 269, 8280-8286.
26. Weidman, C.A., Netter, R.C., Benjamin, L.R., McAllister, J.J., Schmiedekamp, L.A., Coleman, R.A., and Pugh, B.F. (1997) *J. Mol. Biol.* 271, 61-75.
27. Meisterernst, M. and Roeder, R.G. (1991) *Cell* 67, 557-567.
28. Inostroza, J.A., Mermelstein, F.H., Ha, I., Lane, W.S., and Reinberg, D. (1992) *Cell* 70, 477-489.
29. Merino, A., Madden, K.R., Lane, W., Champoux, J.J., and Reinberg, D. (1993) *Nature* 365, 227-232.
30. Ge, H. and Roeder, R.G. (1994) *J. Biol. Chem.* 269, 17136-17140.
31. Kirov, N.C., Lieberman, P.M., and Rushlow, C. (1996) *EMBO J.* 15, 7079-7087.
32. Auble, D.T. and Hahn, S. (1993) *Genes Dev.* 7, 844-856.
33. Auble, D.T., Hansen, K.E., Mueller, C.G., Lane, W.S., Thorner, J., and Hahn, S. (1994) *Genes Dev.* 8, 1920-1934.
34. Chicca, J.J. II, Auble, D.T., and Pugh, B.F. (1998) *Mol. Cell. Biol.* 18, 1701-1710.

5

26. Kokubo, T., Swanson, M.J., Nishikawa, J.I., Hinnebusch, A.G., and Nakatani, Y. (1998) *Mol. Cell. Biol.* 18, 1003-1012.

27. Ozer, J., Mitsouras, K., Zerby, D., Carey, M., and Lieberman, P.M. (1998) *J. Biol. Chem.* 273, 14293-14300.

28. Wang, W., Gralla, J.D., and Carey, M. (1992) *Genes Dev.* 6, 1716-1727.

29. Lieberman, P.M. and Berk, A.J. (1994) *Genes Dev.* 8, 995-1006.

30. Ma, D., Olave, I., Merino, A., and Reinberg, D. (1996) *Proc. Natl. Acad. Sci.* 93, 6583-6588.

10

31. Ozer, J., Bolden, A.H., and Lieberman, P.M. (1996) *J. Biol. Chem.* 271, 11182-11190.

32. Ge, H. and Roeder, R.G. (1994) *Cell* 78, 513-523.

33. Shykind, B.M., Kim, J., and Sharp, P.A. (1995) *Genes Dev.* 9, 1354-1365.

34. Oelgeschlager, T., Chiang, C.-M., and Roeder, R.G. (1996) *Nature* 382, 735-738.

15

35. Chi, T. and Carey, M. (1996) *Genes Dev.* 10, 2540-2550.

36. Emami, K.H., Jain, A., and Smale, S.T. (1997) *Genes Dev.* 11, 3007-3019.

37. Lennon, C.G., Auffray, C., Polymeropoulos, M., and Soares, M.B. (1996) *Genomics* 33, 151-152.

20

38. Kozak, M. (1984) *Nuc. Acids Res.* 12, 857-872.

39. Andrews, J., Smith, M., Merakovsky, J., Coulson, M., Hannan, F., and Kelly, L.E. (1996) *Genetics* 143, 1699-1711.

40. Wilson et al. (1994) *Nature* 368, 32-38.

25

41. Grigliatti, T.A., Hall, L., Rosenbluth, R., and Suzuki, D.T. (1973) *Mol. Gen. Genet.* 120, 107-114.

42. Thurieau, C., Brosius, J., Burne, C., Jolles, P., Keen, J.H., Mattalia, R.J., Chow, E.P., Ramachandran, K.L., and Kirchhausen, T. (1988) *DNA* 7, 663-669.

43. Nakayama, Y., Goebl, M., O'Brine Greco, B., Lemmon, S., Chow, E.P.-C., and Kirchhausen, T. (1991) *Eur. J. Bioch.* 202, 569-574.

25

44. Keen, J.H. (1990) *Ann. Rev. Bioch.* 59, 415-438.

30

45. Hirst, J. and Robinson, M.S. (1998) *Bioch. Biophys. Acta* 1404, 173-193.

What is claimed is:

1. An isolated polynucleotide sequence comprising a nucleotide sequence that encodes a polypeptide of SEQ ID NO. 2, or fragments of said sequence.

5

2. A composition comprising a polypeptide sequence set out in SEQ ID NO. 2 and fragments thereof, or those which are functionally equivalent.

10

3. The DNA segment of claim 1, that encodes the human testis-specific transcriptional factor ALF.

15

4. The DNA segment of claim 1, comprising a human testis-specific transcriptional factor that includes at least 30 contiguous nucleotides from position 16 to 1617 of SEQ ID NO.:1.

20

5. The DNA segment of claim 1, comprising a human testis-specific transcriptional factor that encodes the amino acids of SEQ ID NO.: 2.

6. The DNA segment of claim 1, whereby the segment encodes amino acids 1 to 478 of SEQ ID NO.: 2.

7. The DNA segment of claim 1, positioned under the control of a promoter.

25

8. The DNA segment of claim 1, further defined as comprising a recombinant vector.

9. An isolated nucleic acid segment characterized as:

(a) a nucleic acid segment comprising a sequence that consists essentially of at least 30 contiguous nucleotides that have substantially the same sequence as, or are complementary to, 30 contiguous nucleic acids of SEQ ID NO.:1; or

5 (b) a nucleic acid segment of from 14 to about 1434 nucleotides in length that hybridize to the nucleic acid segment of SEQ ID NO.:1, or complement thereof, under high stringency hybridization conditions.

10 10. The nucleic acid segment of claim 9, wherein the segment comprises a sequence region of at least 30 contiguous nucleotides from SEQ ID NO.:1, or the complement thereof.

15 11. The nucleic acid segment of claim 9, wherein the segment comprises a sequence region of at least about 30 nucleotides; or wherein the segment is about 30 nucleotides in length.

20 12. The nucleic acid segment of claim 11, wherein the segment comprises a sequence region of at least about 30 nucleotides; or wherein the segment is about 30 nucleotides in length.

25 13. The nucleic acid segment of claim 12, wherein the segment comprises a sequence region of at least about 50 nucleotides; or wherein the segment is about 50 nucleotides in length.

14. The nucleic acid segment of claim 13, wherein the segment comprises a sequence region of at least about 100 nucleotides; or wherein the segment is about 100 nucleotides in length.

15. The nucleic acid segment of claim 14, wherein the segment comprises a sequence region of at least about 200 nucleotides; or wherein the segment is about 200 nucleotides in length.

5 16. The nucleic acid segment of claim 15, wherein the segment comprises a sequence region of at least about 500 nucleotides; or wherein the segment is about 500 nucleotides in length.

10 17. The nucleic acid segment of claim 16, wherein the segment comprises a sequence region of at least about 1000 nucleotides; or wherein the segment is about 1000 nucleotides in length.

15 18. The nucleic acid segment of claim 16, wherein the segment comprises a sequence region of about 1434 nucleotides; or wherein the segment is about 1434 nucleotides in length.

20 19. The nucleic acid segment of claim 9, wherein the segment is up to 10,000 base pairs in length.

20 20. The nucleic acid segment of claim 19, wherein the segment is up to 5,000 base pairs in length.

25 21. The nucleic acid segment of claim 20, wherein the segment is up to 3,000 base pairs in length.

22. The nucleic acid segment of claim 21, wherein the segment is up to 1,000 base pairs in length.

23. The nucleic acid segment of claim 9, further defined as a DNA segment.

24. The nucleic acid segment of claim 9, further defined as a RNA segment.

25. A recombinant host cell comprising a DNA segment as defined in claim 9.

5

26. The recombinant host cell of claim 25, wherein the DNA segment is introduced into the cell by means of a recombinant vector.

10 27. The recombinant host cell of claim 25, wherein the whole cell expresses the DNA segment to produce the encoded human testis-specific transcriptional factor protein or peptide wherein said protein or peptide has the amino acid sequence of SEQ ID NO.:2.

28. The recombinant host cell of claim 25, further defined as a bacterial host cell.

15 29. The recombinant host cell of claim 28, wherein the bacterial host cell is E. coli.

30. A method of using a DNA segment that includes an isolated testis-specific transcriptional factor gene, comprising the steps of:

- (a) preparing a recombinant vector in which a testis-specific transcriptional factor gene encoding DNA segment, or a fragment thereof, is positioned under the control of a promoter wherein said testis-specific transcriptional factor gene has the sequence of SEQ ID NO.:2;
- (b) introducing said recombinant vector into recombinant host cell;
- (c) culturing a recombinant host cell under conditions effective to allow expression of the encoded testis-specific transcriptional factor gene protein or peptide; and
- (d) collecting said testis-specific transcriptional factor gene protein or peptide.

10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
140  
145  
150  
155  
160  
165  
170  
175  
180  
185  
190  
195  
200  
205  
210  
215  
220  
225  
230  
235  
240  
245  
250  
255  
260  
265  
270  
275  
280  
285  
290  
295  
300  
305  
310  
315  
320  
325  
330  
335  
340  
345  
350  
355  
360  
365  
370  
375  
380  
385  
390  
395  
400  
405  
410  
415  
420  
425  
430  
435  
440  
445  
450  
455  
460  
465  
470  
475  
480  
485  
490  
495  
500  
505  
510  
515  
520  
525  
530  
535  
540  
545  
550  
555  
560  
565  
570  
575  
580  
585  
590  
595  
600  
605  
610  
615  
620  
625  
630  
635  
640  
645  
650  
655  
660  
665  
670  
675  
680  
685  
690  
695  
700  
705  
710  
715  
720  
725  
730  
735  
740  
745  
750  
755  
760  
765  
770  
775  
780  
785  
790  
795  
800  
805  
810  
815  
820  
825  
830  
835  
840  
845  
850  
855  
860  
865  
870  
875  
880  
885  
890  
895  
900  
905  
910  
915  
920  
925  
930  
935  
940  
945  
950  
955  
960  
965  
970  
975  
980  
985  
990  
995  
1000  
1005  
1010  
1015  
1020  
1025  
1030  
1035  
1040  
1045  
1050  
1055  
1060  
1065  
1070  
1075  
1080  
1085  
1090  
1095  
1100  
1105  
1110  
1115  
1120  
1125  
1130  
1135  
1140  
1145  
1150  
1155  
1160  
1165  
1170  
1175  
1180  
1185  
1190  
1195  
1200  
1205  
1210  
1215  
1220  
1225  
1230  
1235  
1240  
1245  
1250  
1255  
1260  
1265  
1270  
1275  
1280  
1285  
1290  
1295  
1300  
1305  
1310  
1315  
1320  
1325  
1330  
1335  
1340  
1345  
1350  
1355  
1360  
1365  
1370  
1375  
1380  
1385  
1390  
1395  
1400  
1405  
1410  
1415  
1420  
1425  
1430  
1435  
1440  
1445  
1450  
1455  
1460  
1465  
1470  
1475  
1480  
1485  
1490  
1495  
1500  
1505  
1510  
1515  
1520  
1525  
1530  
1535  
1540  
1545  
1550  
1555  
1560  
1565  
1570  
1575  
1580  
1585  
1590  
1595  
1600  
1605  
1610  
1615  
1620  
1625  
1630  
1635  
1640  
1645  
1650  
1655  
1660  
1665  
1670  
1675  
1680  
1685  
1690  
1695  
1700  
1705  
1710  
1715  
1720  
1725  
1730  
1735  
1740  
1745  
1750  
1755  
1760  
1765  
1770  
1775  
1780  
1785  
1790  
1795  
1800  
1805  
1810  
1815  
1820  
1825  
1830  
1835  
1840  
1845  
1850  
1855  
1860  
1865  
1870  
1875  
1880  
1885  
1890  
1895  
1900  
1905  
1910  
1915  
1920  
1925  
1930  
1935  
1940  
1945  
1950  
1955  
1960  
1965  
1970  
1975  
1980  
1985  
1990  
1995  
2000  
2005  
2010  
2015  
2020  
2025  
2030  
2035  
2040  
2045  
2050  
2055  
2060  
2065  
2070  
2075  
2080  
2085  
2090  
2095  
2100  
2105  
2110  
2115  
2120  
2125  
2130  
2135  
2140  
2145  
2150  
2155  
2160  
2165  
2170  
2175  
2180  
2185  
2190  
2195  
2200  
2205  
2210  
2215  
2220  
2225  
2230  
2235  
2240  
2245  
2250  
2255  
2260  
2265  
2270  
2275  
2280  
2285  
2290  
2295  
2300  
2305  
2310  
2315  
2320  
2325  
2330  
2335  
2340  
2345  
2350  
2355  
2360  
2365  
2370  
2375  
2380  
2385  
2390  
2395  
2400  
2405  
2410  
2415  
2420  
2425  
2430  
2435  
2440  
2445  
2450  
2455  
2460  
2465  
2470  
2475  
2480  
2485  
2490  
2495  
2500  
2505  
2510  
2515  
2520  
2525  
2530  
2535  
2540  
2545  
2550  
2555  
2560  
2565  
2570  
2575  
2580  
2585  
2590  
2595  
2600  
2605  
2610  
2615  
2620  
2625  
2630  
2635  
2640  
2645  
2650  
2655  
2660  
2665  
2670  
2675  
2680  
2685  
2690  
2695  
2700  
2705  
2710  
2715  
2720  
2725  
2730  
2735  
2740  
2745  
2750  
2755  
2760  
2765  
2770  
2775  
2780  
2785  
2790  
2795  
2800  
2805  
2810  
2815  
2820  
2825  
2830  
2835  
2840  
2845  
2850  
2855  
2860  
2865  
2870  
2875  
2880  
2885  
2890  
2895  
2900  
2905  
2910  
2915  
2920  
2925  
2930  
2935  
2940  
2945  
2950  
2955  
2960  
2965  
2970  
2975  
2980  
2985  
2990  
2995  
3000  
3005  
3010  
3015  
3020  
3025  
3030  
3035  
3040  
3045  
3050  
3055  
3060  
3065  
3070  
3075  
3080  
3085  
3090  
3095  
3100  
3105  
3110  
3115  
3120  
3125  
3130  
3135  
3140  
3145  
3150  
3155  
3160  
3165  
3170  
3175  
3180  
3185  
3190  
3195  
3200  
3205  
3210  
3215  
3220  
3225  
3230  
3235  
3240  
3245  
3250  
3255  
3260  
3265  
3270  
3275  
3280  
3285  
3290  
3295  
3300  
3305  
3310  
3315  
3320  
3325  
3330  
3335  
3340  
3345  
3350  
3355  
3360  
3365  
3370  
3375  
3380  
3385  
3390  
3395  
3400  
3405  
3410  
3415  
3420  
3425  
3430  
3435  
3440  
3445  
3450  
3455  
3460  
3465  
3470  
3475  
3480  
3485  
3490  
3495  
3500  
3505  
3510  
3515  
3520  
3525  
3530  
3535  
3540  
3545  
3550  
3555  
3560  
3565  
3570  
3575  
3580  
3585  
3590  
3595  
3600  
3605  
3610  
3615  
3620  
3625  
3630  
3635  
3640  
3645  
3650  
3655  
3660  
3665  
3670  
3675  
3680  
3685  
3690  
3695  
3700  
3705  
3710  
3715  
3720  
3725  
3730  
3735  
3740  
3745  
3750  
3755  
3760  
3765  
3770  
3775  
3780  
3785  
3790  
3795  
3800  
3805  
3810  
3815  
3820  
3825  
3830  
3835  
3840  
3845  
3850  
3855  
3860  
3865  
3870  
3875  
3880  
3885  
3890  
3895  
3900  
3905  
3910  
3915  
3920  
3925  
3930  
3935  
3940  
3945  
3950  
3955  
3960  
3965  
3970  
3975  
3980  
3985  
3990  
3995  
4000  
4005  
4010  
4015  
4020  
4025  
4030  
4035  
4040  
4045  
4050  
4055  
4060  
4065  
4070  
4075  
4080  
4085  
4090  
4095  
4100  
4105  
4110  
4115  
4120  
4125  
4130  
4135  
4140  
4145  
4150  
4155  
4160  
4165  
4170  
4175  
4180  
4185  
4190  
4195  
4200  
4205  
4210  
4215  
4220  
4225  
4230  
4235  
4240  
4245  
4250  
4255  
4260  
4265  
4270  
4275  
4280  
4285  
4290  
4295  
4300  
4305  
4310  
4315  
4320  
4325  
4330  
4335  
4340  
4345  
4350  
4355  
4360  
4365  
4370  
4375  
4380  
4385  
4390  
4395  
4400  
4405  
4410  
4415  
4420  
4425  
4430  
4435  
4440  
4445  
4450  
4455  
4460  
4465  
4470  
4475  
4480  
4485  
4490  
4495  
4500  
4505  
4510  
4515  
4520  
4525  
4530  
4535  
4540  
4545  
4550  
4555  
4560  
4565  
4570  
4575  
4580  
4585  
4590  
4595  
4600  
4605  
4610  
4615  
4620  
4625  
4630  
4635  
4640  
4645  
4650  
4655  
4660  
4665  
4670  
4675  
4680  
4685  
4690  
4695  
4700  
4705  
4710  
4715  
4720  
4725  
4730  
4735  
4740  
4745  
4750  
4755  
4760  
4765  
4770  
4775  
4780  
4785  
4790  
4795  
4800  
4805  
4810  
4815  
4820  
4825  
4830  
4835  
4840  
4845  
4850  
4855  
4860  
4865  
4870  
4875  
4880  
4885  
4890  
4895  
4900  
4905  
4910  
4915  
4920  
4925  
4930  
4935  
4940  
4945  
4950  
4955  
4960  
4965  
4970  
4975  
4980  
4985  
4990  
4995  
5000  
5005  
5010  
5015  
5020  
5025  
5030  
5035  
5040  
5045  
5050  
5055  
5060  
5065  
5070  
5075  
5080  
5085  
5090  
5095  
5100  
5105  
5110  
5115  
5120  
5125  
5130  
5135  
5140  
5145  
5150  
5155  
5160  
5165  
5170  
5175  
5180  
5185  
5190  
5195  
5200  
5205  
5210  
5215  
5220  
5225  
5230  
5235  
5240  
5245  
5250  
5255  
5260  
5265  
5270  
5275  
5280  
5285  
5290  
5295  
5300  
5305  
5310  
5315  
5320  
5325  
5330  
5335  
5340  
5345  
5350  
5355  
5360  
5365  
5370  
5375  
5380  
5385  
5390  
5395  
5400  
5405  
5410  
5415  
5420  
5425  
5430  
5435  
5440  
5445  
5450  
5455  
5460  
5465  
5470  
5475  
5480  
5485  
5490  
5495  
5500  
5505  
5510  
5515  
5520  
5525  
5530  
5535  
5540  
5545  
5550  
5555  
5560  
5565  
5570  
5575  
5580  
5585  
5590  
5595  
5600  
5605  
5610  
5615  
5620  
5625  
5630  
5635  
5640  
5645  
5650  
5655  
5660  
5665  
5670  
5675  
5680  
5685  
5690  
5695  
5700  
5705  
5710  
5715  
5720  
5725  
5730  
5735  
5740  
5745  
5750  
5755  
5760  
5765  
5770  
5775  
5780  
5785  
5790  
5795  
5800  
5805  
5810  
5815  
5820  
5825  
5830  
5835  
5840  
5845  
5850  
5855  
5860  
5865  
5870  
5875  
5880  
5885  
5890  
5895  
5900  
5905  
5910  
5915  
5920  
5925  
5930  
5935  
5940  
5945  
5950  
5955  
5960  
5965  
5970  
5975  
5980  
5985  
5990  
5995  
6000  
6005  
6010  
6015  
6020  
6025  
6030  
6035  
6040  
6045  
6050  
6055  
6060  
6065  
6070  
6075  
6080  
6085  
6090  
6095  
6100  
6105  
6110  
6115  
6120  
6125  
6130  
6135  
6140  
6145  
6150  
6155  
6160  
6165  
6170  
6175  
6180  
6185  
6190  
6195  
6200  
6205  
6210  
6215  
6220  
6225  
6230  
6235  
6240  
6245  
6250  
6255  
6260  
6265  
6270  
6275  
6280  
6285  
6290  
6295  
6300  
6305  
6310  
6315  
6320  
6325  
6330  
6335  
6340  
6345  
6350  
6355  
6360  
6365  
6370  
6375  
6380  
6385  
6390  
6395  
6400  
6405  
6410  
6415  
6420  
6425  
6430  
6435  
6440  
6445  
6450  
6455  
6460  
6465  
6470  
6475  
6480  
6485  
6490  
6495  
6500  
6505  
6510  
6515  
6520  
6525  
6530  
6535  
6540  
6545  
6550  
6555  
6560  
6565  
6570  
6575  
6580  
6585  
6590  
6595  
6600  
6605  
6610  
6615  
6620  
6625  
6630  
6635  
6640  
6645  
6650  
6655  
6660  
6665  
6670  
6675  
6680  
6685  
6690  
6695  
6700  
6705  
6710  
6715  
6720  
6725  
6730  
6735  
6740  
6745  
6750  
6755  
6760  
6765  
6770  
6775  
6780  
6785  
6790  
6795  
6800  
6805  
6810  
6815  
6820  
6825  
6830  
6835  
6840  
6845  
6850  
6855  
6860  
6865  
6870  
6875  
6880  
6885  
6890  
6895  
6900  
6905  
6910  
6915  
6920  
6925  
6930  
6935  
6940  
6945  
6950  
6955  
6960  
6965  
6970  
6975  
6980  
6985  
6990  
6995  
7000  
7005  
7010  
7015  
7020  
7025  
7030  
7035  
7040  
7045  
7050  
7055  
7060  
7065  
7070  
7075  
7080  
7085  
7090  
7095  
7100  
7105  
7110  
7115  
7120  
7125  
7130  
7135  
7140  
7145  
7150  
7155  
7160  
7165  
7170  
7175  
7180  
7185  
7190  
7195  
7200  
7205  
7210  
7215  
7220  
7225  
7230  
7235  
7240  
7245  
7250  
7255  
7260  
7265  
7270  
7275  
7280  
7285  
7290  
7295  
7300  
7305  
7310  
7315  
7320  
7325  
7330  
7335  
7340  
7345  
7350  
7355  
7360  
7365  
7370  
7375  
7380  
7385  
7390  
7395  
7400  
7405  
7410  
7415  
7420  
7425  
7430  
7435  
7440  
7445  
7450  
7455  
7460  
7465  
7470  
7475  
7480  
7485  
7490  
7495  
7500  
7505  
7510  
7515  
7520  
7525  
7530  
7535  
7540  
7545  
7550  
7555  
7560  
7565  
7570  
7575  
7580  
7585  
7590  
7595  
7600  
7605  
7610  
7615  
7620  
7625  
7630  
7635  
7640  
7645  
7650  
7655  
7660  
7665  
7670  
7675  
7680  
7685  
7690  
7695  
7700  
7705  
7710  
7715  
7720  
7725  
7730  
7735  
7740  
7745  
7750  
7755  
7760  
7765  
7770  
7775  
7780  
7785  
7790  
7795  
7800  
7805  
7810  
7815  
7820  
7825  
7830  
7835  
7840  
7845  
7850  
7855  
7860  
7865  
7870  
7875  
7880  
7885  
7890  
7895  
7900  
7905  
7910  
7915  
7920  
7925  
7930  
7935  
7940  
7945  
7950  
7955  
7960  
7965  
7970  
7975  
7980  
7985  
7990  
7995  
8000  
8005  
8010  
8015  
8020  
8025  
8030  
8035  
8040  
8045  
8050  
8055  
8060  
8065  
8070  
8075  
8080  
8085  
8090  
8095  
8100  
8105  
8110  
8115  
8120  
8125  
8130  
8135  
8140  
8145  
8150  
8155  
8160  
8165  
8170  
8175  
8180  
8185  
8190  
8195  
8200  
8205  
8210  
8215  
8220  
8225  
8230  
8235  
8240  
8245  
8250  
8255  
8260  
8265  
8270  
8275  
8280  
8285  
8290  
8295  
8300  
8305  
8310  
8315  
8320  
8325  
8330  
8335  
8340  
8345  
8350  
8355  
8360  
8365  
8370  
8375  
8380  
8385  
8390  
8395  
8400  
8405  
8410  
8415  
8420  
8425  
8430  
8435  
8440  
8445  
8450  
8455  
8460  
8465  
8470  
8475  
8480  
8485  
8490  
8495  
8500  
8505  
8510  
8515  
8520  
8525  
8530  
8535  
8540  
8545  
8550  
8555  
8560  
8565  
8570  
8575  
8580  
8585  
8590  
8595  
8600  
8605  
8610  
8615  
8620  
8625  
8630  
8635  
8640  
8645  
8650  
8655  
8660  
8665  
8670  
8675  
8680  
8685  
8690  
8695  
8700  
8705  
8710  
8715  
8720  
8725  
8730  
8735  
8740  
8745  
8750  
8755  
8760  
8765  
8770  
8775  
8780  
8785  
8790  
8795  
8800  
8805  
8810  
8815  
8820  
8825  
8830  
8835  
8840  
8845  
8850  
8855  
8860  
8865  
8870  
8875  
8880  
8885  
8890  
8895  
8900  
8905  
8910  
8915  
8920  
8925  
8930  
8935  
8940  
8945  
8950  
8955  
8960  
8965  
8970  
8975  
8980  
8985  
8990  
8995  
9000  
9005  
9010  
9015  
9020  
9025  
9030  
9035  
9040  
9045  
9050  
9055  
9060  
9065  
9070  
9075  
9080  
9085  
9090  
9095  
9100  
9105  
9110  
9115  
9120  
9125  
9130  
9135  
9140  
9145  
9150  
9155  
9160  
9165  
9170  
9175  
9180  
9185  
9190  
9195  
9200  
9205  
9210  
9215  
9220  
9225  
9230  
9235  
9240  
9245  
9250  
9255  
9260  
9265  
9270  
9275  
9280  
9285  
9290  
9295  
9300  
9305  
9310  
9315  
9320  
9325  
9330  
9335  
9340  
9345  
9350  
9355  
9360  
9365  
9370  
9375  
9380  
9385  
9390  
9395  
9400  
9405  
9410  
9415  
9420  
9425  
9430  
9435  
9440  
9445  
9450  
9455  
9460  
946

36. The isolated polynucleotide sequence of claim 35, comprising a nucleic acid molecule that comprises the nucleotide sequence of SEQ ID NO. 3, or fragments of said sequence.

5 37. The DNA segment of claim 35, that encodes a human transcriptional factor related to the *Drosophila* Stoned gene.

10 38. The DNA segment of claim 35, that encodes the human transcriptional factor SALF.

15 39. The DNA segment of claim 35, comprising a human testis-specific transcriptional factor that includes at least 30 contiguous nucleotides from position 16 to 3824 of SEQ ID NO.:3.

20 40. The DNA segment of claim 35, comprising a human testis-specific transcriptional factor that encodes the amino acids of SEQ ID NO.: 4.

25 41. The DNA segment of claim 35, which encodes amino acids 1 to 11821 of SEQ ID NO.: 4.

42. The DNA segment of claim 35, positioned under the control of a promoter.

43. The DNA segment of claim 35, further defined as comprising a recombinant vector.

44. An isolated nucleic acid segment characterized as:

(a) a nucleic acid segment comprising a sequence that consists essentially of at least 30 contiguous nucleotides that have substantially the same sequence as, or are complementary to, 30 contiguous nucleic acids of SEQ ID NO.:3; or

(b) a nucleic acid segment of from 14 to about 3824 nucleotides in length that hybridize to the nucleic acid segment of SEQ ID NO.:3, or complement thereof, under high stringency hybridization conditions.

5        45. The nucleic acid segment of claim 43, wherein the segment comprises a sequence region of at least 30 contiguous nucleotides from SEQ ID NO.:3, or the complement thereof.

46. The nucleic acid segment of claim 43, wherein the segment comprises a  
sequence region of at least about 30 nucleotides; or wherein the segment is about 30  
nucleotides in length.

47. The nucleic acid segment of claim 45, wherein the segment comprises a sequence region of at least about 30 nucleotides; or wherein the segment is about 30 nucleotides in length.

48. The nucleic acid segment of claim 46, wherein the segment comprises a sequence region of at least about 50 nucleotides; or wherein the segment is about 50 nucleotides in length.

49. The nucleic acid segment of claim 47, wherein the segment comprises a sequence region of at least about 100 nucleotides; or wherein the segment is about 100 nucleotides in length.

25        50. The nucleic acid segment of claim 48, wherein the segment comprises a sequence region of at least about 200 nucleotides; or wherein the segment is about 200 nucleotides in length.

51. The nucleic acid segment of claim 49, wherein the segment comprises a sequence region of at least about 500 nucleotides; or wherein the segment is about 500 nucleotides in length.

5 52. The nucleic acid segment of claim 50, wherein the segment comprises a sequence region of at least about 1000 nucleotides; or wherein the segment is about 1000 nucleotides in length.

10 53. The nucleic acid segment of claim 51, wherein the segment comprises a sequence region of about 3824 nucleotides; or wherein the segment is about 3824 nucleotides in length.

15 54. The nucleic acid segment of claim 43, wherein the segment is up to 10,000 base pairs in length.

20 55. The nucleic acid segment of claim 53, wherein the segment is up to 5,000 base pairs in length.

56. The nucleic acid segment of claim 54, wherein the segment is up to 3,000 base pairs in length.

25 57. The nucleic acid segment of claim 55, wherein the segment is up to 1,000 base pairs in length.

58. The nucleic acid segment of claim 43, further defined as a DNA segment.

59. The nucleic acid segment of claim 43, further defined as a RNA segment.

60. A recombinant host cell comprising a DNA segment as defined in claim 43.

5           61. The recombinant host cell of claim 59, wherein the DNA segment is introduced into the cell by means of a recombinant vector.

10          62. The recombinant host cell of claim 59, wherein the whole cell expresses the DNA segment to produce the encoded transcriptional factor protein or peptide wherein said protein or peptide has the amino acid sequence of SEQ ID NO.:4.

15          63. The recombinant host cell of claim 59, further defined as a bacterial host cell.

20          64. The recombinant host cell of claim 62, wherein the bacterial host cell is E. coli.

25          65. A method of using a DNA segment that includes an isolated testis-specific transcriptional factor gene, comprising the steps of:

- (a) preparing a recombinant vector in which a human transcriptional factor gene encoding DNA segment is positioned under the control of a promoter wherein said testis-specific transcriptional factor gene has the sequence of SEQ ID NO.:4;
- (b) introducing said recombinant vector into recombinant host cell;
- (c) culturing a recombinant host cell under conditions effective to allow expression of the encoded transcriptional factor gene protein or peptide; and
- (d) collecting the human testis-specific transcriptional factor gene protein or peptide.

25          66. A method of making a recombinant vector comprising inserting the isolated DNA segment of SEQ ID NO.:3 into a vector.

67. An isolated DNA segment comprising the sequence of Genbank Accession number AF026169, its complement, and fragments of said sequence.

5       68. An isolated polypeptide comprising a mature polypeptide having an amino acid sequence encoded by a nucleic acid segment that is at least 95% identical to SEQ ID NO.:3.

10      69. The isolated polypeptide of claim 20, comprising amino acids 1 to 1182 of SEQ ID NO.: 4.

15      70. An isolated nucleic acid segment characterized as:

- (a) a nucleic acid segment comprising a sequence that consists essentially of at least 30 contiguous nucleotides that have substantially the same sequence as, or are complementary to, 30 contiguous nucleic acids of SEQ ID NO.:5; or
- (b) a nucleic acid segment of from 1 to about 30 nucleotides in length that hybridize to the nucleic acid segment of SEQ ID NO.:5, or complement thereof, under high stringency hybridization conditions.

GENBANK AF026169

**TRANSCRIPTION FACTORS RELATED TO TFIIA**

**ABSTRACT OF THE DISCLOSURE**

The invention provides human polynucleotide sequences that encode transcription factor polypeptides that are termed ALF and SALF, and an alternative C-terminal sequence. The invention includes ALF, SALF and alternative C-terminus polypeptides, peptides, fusion proteins, expression vectors, agonists, antagonists, host cells that overexpress these polypeptides, including transgenic animals, and recombinant knock-out animals that cannot express the relevant RNAs and polypeptides. The invention also provides methods for the detection, diagnosis, screening, and monitoring disorders related to inappropriate expression, production, or activity of ALF and SALF, and provides methods to increase or decrease gene expression with respect to treating disorders related to inappropriate or ineffectual patterns of gene expression.

691850.4

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION: Jeff L. DeJong )  
SERIAL NO.: )  
FILED: Concurrently Herewith )  
FOR: TRANSCRIPTION FACTORS )  
RELATED TO TFIIA )

Box Patent Application  
Assistant Commissioner  
for Patents  
Washington, D.C. 20231

"EXPRESS MAILING" Mailing Label No.  
EL152179982US. Date of Deposit: June 4, 1999. I  
hereby certify that this paper is being deposited with  
the U.S. Postal Service Express Mail Post Office to  
Addressee Service under 37 CFR 1.10 on the date  
shown above and is addressed to the Ass't  
Commissioner for Patents, Washington, D.C. 20231.

  
Rita J. Carr

SEQUENCE LISTING

Dear Sir:

Kindly accept the attached Sequence Listing and disk related to the above-identified  
Application for Patent.

Attached please find a hard copy of a Sequence Listing along with a diskette containing a  
Sequence Listing in computer readable form in ASCII-DOS format in accordance with the  
requirements of 37 C.F.R. 1.824.

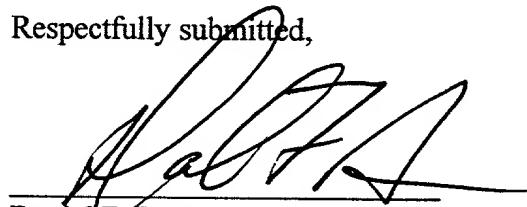
REMARKS

A diskette with a Sequence Listing in ASCII-DOS format is attached and consideration of the application is requested. The content of the paper and computer readable copies are the same as required by 37 C.F.R. 1.821(f).

For the foregoing reasons, the Sequence Listing is submitted to be in order, and complies with the Sequence Rules 37 C.F.R. 1.821 - 1.825.

Dated this 4<sup>th</sup> day of June, 1999.

Respectfully submitted,



Daniel E. Perez  
Attorney for Applicant  
Reg. No. 33,755  
Gardere & Wynne, L.L.P.  
1601 Elm Street, Suite 3000  
Dallas, TX 75201  
(214) 999-4355 - Tel  
(214) 999-4667 - Fax

730212.1

SEQUENCE LISTING

<110> DeJong Dr., Jeff

<120> Transcription Factors Related to TFIIA

<130> 119941-1053

<140>

<141>

<160> 19

<170> PatentIn Ver. 2.0

<210> 1

<211> 1617

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (16)..(1449)

<400> 1

gctggaggtg ctgtc atg gcc tgc ctc aac ccg gtg cct aaa ctc tac aga 51

Met Ala Cys Leu Asn Pro Val Pro Lys Leu Tyr Arg

1

5

10

tct gta att gaa gat gta att gaa gga gtt cgg aat cta ttt gct gaa 99

Ser Val Ile Glu Asp Val Ile Glu Gly Val Arg Asn Leu Phe Ala Glu

15

20

25

gaa ggt ata gag gaa caa gtt tta aaa gac ttg aag cag ctc tgg gaa 147

Glu Gly Ile Glu Glu Gln Val Leu Lys Asp Leu Lys Gln Leu Trp Glu

30 35 40

acc aag gtt ttg cag tct aaa gca aca gaa gac ttc ttc aga aat agc 195

Thr Lys Val Leu Gln Ser Lys Ala Thr Glu Asp Phe Phe Arg Asn Ser

45 50 55 60

atc caa tca cct ctg ttt act ctt cag ttg ccg cac agc ttg cac caa 243

Ile Gln Ser Pro Leu Phe Thr Leu Gln Leu Pro His Ser Leu His Gln

65 70 75

aca ttg caa tcg tca aca gca tca tta gtt att cct gct ggt aga act 291

Thr Leu Gln Ser Ser Thr Ala Ser Leu Val Ile Pro Ala Gly Arg Thr

80 85 90

ctt cca agt ttt acc aca gca gaa ctg ggc act tca aac tcc agt gca 339

Leu Pro Ser Phe Thr Thr Ala Glu Leu Gly Thr Ser Asn Ser Ser Ala

95 100 105

aac ttt act ttt cct ggt tat ccc att cat gta cca gca ggt gtg aca 387

Asn Phe Thr Phe Pro Gly Tyr Pro Ile His Val Pro Ala Gly Val Thr

110 115 120

cta cag act gta tct ggt cac ctt tat aaa gtc aat gta cca att atg 435

Leu Gln Thr Val Ser Gly His Leu Tyr Lys Val Asn Val Pro Ile Met

125 130 135 140

gtg aca gag act tct gga aga gca ggt att ctt cag cat cca att cag 483

Val Thr Glu Thr Ser Gly Arg Ala Gly Ile Leu Gln His Pro Ile Gln

145 150 155

caa gta ttt caa cag ctt ggc cag cct tca gta ata caa act agt gtt 531

Gln Val Phe Gln Gln Leu Gly Gln Pro Ser Val Ile Gln Thr Ser Val

160 165 170

cca caa ttg aat cca tgg tct ctt caa gca act act gaa aaa tca cag 579  
Pro Gln Leu Asn Pro Trp Ser Leu Gln Ala Thr Thr Glu Lys Ser Gln  
175 180 185

aga att gaa acc gtg cta cag caa ccc gca att cta cct tct ggg cca 627  
Arg Ile Glu Thr Val Leu Gln Gln Pro Ala Ile Leu Pro Ser Gly Pro  
190 195 200

gta gat agg aaa cac tta gaa aat gcc acc agt gat ata ctt gta tct 675  
Val Asp Arg Lys His Leu Glu Asn Ala Thr Ser Asp Ile Leu Val Ser  
205 210 215 220

cct gga aat gag cat aaa atc gtg cct gaa gct ttg ttg tgt cat cag 723  
Pro Gly Asn Glu His Lys Ile Val Pro Glu Ala Leu Leu Cys His Gln  
225 230 235

gaa agt tct cac tat atc agt ctt cca ggt gtt gta ttt tct cca cag 771  
Glu Ser Ser His Tyr Ile Ser Leu Pro Gly Val Val Phe Ser Pro Gln  
240 245 250

gtc tct caa aca aat tct gat gtg gag tca gtg ctc agt ggt tca gct 819  
Val Ser Gln Thr Asn Ser Asp Val Glu Ser Val Leu Ser Gly Ser Ala  
255 260 265

agc atg gct caa aat ctg cat gat gag tcc ctc tcc aca agc cct cat 867  
Ser Met Ala Gln Asn Leu His Asp Glu Ser Leu Ser Thr Ser Pro His  
270 275 280

ggg gct ctc cac cag cac gtg act gat att cag ctt cat att ctt aaa 915  
Gly Ala Leu His Gln His Val Thr Asp Ile Gln Leu His Ile Leu Lys  
285 290 295 300

aat agg atg tat gga tgt gat tct gta aag caa cca aga aat ata gag 963

Asn Arg Met Tyr Gly Cys Asp Ser Val Lys Gln Pro Arg Asn Ile Glu

305

310

315

gaa ccc agc aac ata cct gta tca gag aag gat tct aat tct cag gtg 1011

Glu Pro Ser Asn Ile Pro Val Ser Glu Lys Asp Ser Asn Ser Gln Val

320

325

330

gat tta agc att cgg gtt act gat gat att ggt gaa ata att caa 1059

Asp Leu Ser Ile Arg Val Thr Asp Asp Asp Ile Gly Glu Ile Ile Gln

335

340

345

gta gat gga agc ggt gat aca tct tcc aat gaa gaa ata gga agt aca 1107

Val Asp Gly Ser Gly Asp Thr Ser Ser Asn Glu Glu Ile Gly Ser Thr

350

355

360

aga gat gca gat gag aat gaa ttt cta ggg aat att gac ggg gga gat 1155

Arg Asp Ala Asp Glu Asn Glu Phe Leu Gly Asn Ile Asp Gly Gly Asp

365

370

375

380

ctg aag gta cct gaa gaa gct gac agt att tca aat gag gat tca 1203

Leu Lys Val Pro Glu Glu Ala Asp Ser Ile Ser Asn Glu Asp Ser

385

390

395

gcc aca aac agt agt gat aat gaa gac cct caa gta aac att gta gaa 1251

Ala Thr Asn Ser Ser Asp Asn Glu Asp Pro Gln Val Asn Ile Val Glu

400

405

410

gag gac cct tta aat tct gga gat gat gtt agt gaa cag gat gtg cca 1299

Glu Asp Pro Leu Asn Ser Gly Asp Asp Val Ser Glu Gln Asp Val Pro

415

420

425

gac ctg ttt gac acg gat aat gta att gtc tgt cag tat gat aag att 1347

Asp Leu Phe Asp Thr Asp Asn Val Ile Val Cys Gln Tyr Asp Lys Ile

430

435

440

cat cga agc aag aac aaa tgg aaa ttc tat ttg aaa gat ggt gtt atg 1395  
His Arg Ser Lys Asn Lys Trp Lys Phe Tyr Leu Lys Asp Gly Val Met  
445 450 455 460

tgt ttt gga ggg aga gac tat gta ttt gca aaa gcc att ggt gat gca 1443  
Cys Phe Gly Gly Arg Asp Tyr Val Phe Ala Lys Ala Ile Gly Asp Ala  
465 470 475

gag tgg taaaccttgt gagctcagta catctattt gtgaacatca gttggactat 1499  
Glu Trp

attgcattt gtgaattcat ttttattttt aatatagtc agcacagagc tgttcaaatt 1559

ttagttcac tgtatggaat ttaataaaat tataattcag atgcagatac aattacac 1617

<210> 2  
<211> 478  
<212> PRT  
<213> Homo sapiens

<400> 2  
Met Ala Cys Leu Asn Pro Val Pro Lys Leu Tyr Arg Ser Val Ile Glu  
1 5 10 15

Asp Val Ile Glu Gly Val Arg Asn Leu Phe Ala Glu Glu Gly Ile Glu  
20 25 30

Glu Gln Val Leu Lys Asp Leu Lys Gln Leu Trp Glu Thr Lys Val Leu  
35 40 45

Gln Ser Lys Ala Thr Glu Asp Phe Phe Arg Asn Ser Ile Gln Ser Pro  
50 55 60

Leu Phe Thr Leu Gln Leu Pro His Ser Leu His Gln Thr Leu Gln Ser  
65 70 75 80

Ser Thr Ala Ser Leu Val Ile Pro Ala Gly Arg Thr Leu Pro Ser Phe  
85 90 95

Thr Thr Ala Glu Leu Gly Thr Ser Asn Ser Ala Asn Phe Thr Phe  
100 105 110

Pro Gly Tyr Pro Ile His Val Pro Ala Gly Val Thr Leu Gln Thr Val  
115 120 125

Ser Gly His Leu Tyr Lys Val Asn Val Pro Ile Met Val Thr Glu Thr  
130 135 140

Ser Gly Arg Ala Gly Ile Leu Gln His Pro Ile Gln Gln Val Phe Gln  
145 150 155 160

Gln Leu Gly Gln Pro Ser Val Ile Gln Thr Ser Val Pro Gln Leu Asn  
165 170 175

Pro Trp Ser Leu Gln Ala Thr Thr Glu Lys Ser Gln Arg Ile Glu Thr  
180 185 190

Val Leu Gln Gln Pro Ala Ile Leu Pro Ser Gly Pro Val Asp Arg Lys  
195 200 205

His Leu Glu Asn Ala Thr Ser Asp Ile Leu Val Ser Pro Gly Asn Glu  
210 215 220

His Lys Ile Val Pro Glu Ala Leu Leu Cys His Gln Glu Ser Ser His  
225 230 235 240

Tyr Ile Ser Leu Pro Gly Val Val Phe Ser Pro Gln Val Ser Gln Thr

245 250 255

Asn Ser Asp Val Glu Ser Val Leu Ser Gly Ser Ala Ser Met Ala Gln

260 265 270

Asn Leu His Asp Glu Ser Leu Ser Thr Ser Pro His Gly Ala Leu His

275 280 285

Gln His Val Thr Asp Ile Gln Leu His Ile Leu Lys Asn Arg Met Tyr

290 295 300

Gly Cys Asp Ser Val Lys Gln Pro Arg Asn Ile Glu Glu Pro Ser Asn

305 310 315 320

Ile Pro Val Ser Glu Lys Asp Ser Asn Ser Gln Val Asp Leu Ser Ile

325 330 335

Arg Val Thr Asp Asp Asp Ile Gly Glu Ile Ile Gln Val Asp Gly Ser

340 345 350

Gly Asp Thr Ser Ser Asn Glu Glu Ile Gly Ser Thr Arg Asp Ala Asp

355 360 365

Glu Asn Glu Phe Leu Gly Asn Ile Asp Gly Gly Asp Leu Lys Val Pro

370 375 380

Glu Glu Glu Ala Asp Ser Ile Ser Asn Glu Asp Ser Ala Thr Asn Ser

385 390 395 400

Ser Asp Asn Glu Asp Pro Gln Val Asn Ile Val Glu Glu Asp Pro Leu

405 410 415

Asn Ser Gly Asp Asp Val Ser Glu Gln Asp Val Pro Asp Leu Phe Asp

420

425

430

Thr Asp Asn Val Ile Val Cys Gln Tyr Asp Lys Ile His Arg Ser Lys

435

440

445

Asn Lys Trp Lys Phe Tyr Leu Lys Asp Gly Val Met Cys Phe Gly Gly

450

455

460

Arg Asp Tyr Val Phe Ala Lys Ala Ile Gly Asp Ala Glu Trp

465

470

475

<210> 3

<211> 3824

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (115)..(3660)

<400> 3

ggactttggg actggacaga cctggtcaca gtctagggtc tacatttac tggtcgagca 60

acttttaggt caacctatTTT gatttcttga caagaccaca atctgatccc aaag atg 117

Met

1

tgc tcc aca aat cca ggc aaa tgg gtc acc ttt gat gat gat cct gct 165

Cys Ser Thr Asn Pro Gly Lys Trp Val Thr Phe Asp Asp Asp Pro Ala

5

10

15

gtt caa tct tct caa aag tca aag aat ttt cct ctg gag aat caa ggt 213

Val Gln Ser Ser Gln Lys Ser Asn Phe Pro Leu Glu Asn Gln Gly

20

25

30

gtc tgt aga cca aat gga ctg aag ctg aac cct cct ggc ctc agg gaa 261  
Val Cys Arg Pro Asn Gly Leu Lys Leu Asn Pro Pro Gly Leu Arg Glu

35

40

45

ttt ccc agt gga tct tcc tcc acc agc agc act cct ctc tcc tcc ccc 309  
Phe Pro Ser Gly Ser Ser Ser Thr Ser Ser Thr Pro Leu Ser Ser Pro  
50 55 60 65

att gta gat ttt tat ttc agt cca gga cct cca agt aac tct cct ctt 357  
Ile Val Asp Phe Tyr Phe Ser Pro Gly Pro Pro Ser Asn Ser Pro Leu  
70 75 80

tct aca cct acc aaa gac ttc cca ggt ttt cct ggc atc ccc aaa gca 405  
Ser Thr Pro Thr Lys Asp Phe Pro Gly Phe Pro Gly Ile Pro Lys Ala  
85 90 95

ggg act cat gtg ctt tat cct att cca gaa tca tct tca gac agc cca 453  
Gly Thr His Val Leu Tyr Pro Ile Pro Glu Ser Ser Ser Asp Ser Pro  
100 105 110

ctc gca ata tca gga gga gaa tct tcc tta ctg cct acc aga cca aca 501  
Leu Ala Ile Ser Gly Gly Ser Ser Leu Leu Pro Thr Arg Pro Thr  
115 120 125

tgt tta tcc cat gcc ttg tta ccc agt gac cac tca tgt aca cat cca 549  
Cys Leu Ser His Ala Leu Leu Pro Ser Asp His Ser Cys Thr His Pro  
130 135 140 145

act ccc aaa gta ggt ctt cca gat gaa gtt aat cct caa cag gct gaa 597  
Thr Pro Lys Val Gly Leu Pro Asp Glu Val Asn Pro Gln Gln Ala Glu  
150 155 160

agc cta gga ttc caa agt gat gat ctc ccc cag ttt cag tat ttt cga 645  
Ser Leu Gly Phe Gln Ser Asp Asp Leu Pro Gln Phe Gln Tyr Phe Arg  
165 170 175

gag gac tgt gct ttt tca agt cca ttt cg<sup>g</sup> aaa gat gaa ggc agt gat 693  
Glu Asp Cys Ala Phe Ser Ser Pro Phe Arg Lys Asp Glu Gly Ser Asp  
180 185 190

tcc cat ttc acc ctt gac cca cca gga agc aaa aag atg ttc tca tca 741  
Ser His Phe Thr Leu Asp Pro Pro Gly Ser Lys Lys Met Phe Ser Ser  
195 200 205

aga aac aag gag atg cct att gac caa aaa agc cta aat aag tgt tca 789  
Arg Asn Lys Glu Met Pro Ile Asp Gln Lys Ser Leu Asn Lys Cys Ser  
210 215 220 225

ctc aac tat atc tgt gag aag ctt gaa cat ctc cag tca gct gag aac 837  
Leu Asn Tyr Ile Cys Glu Lys Leu Glu His Leu Gln Ser Ala Glu Asn  
230 235 240

caa gac tca ctt aga agt ttg tct atg cac tgt cta tgt gct gaa gaa 885  
Gln Asp Ser Leu Arg Ser Leu Ser Met His Cys Leu Cys Ala Glu Glu  
245 250 255

aat gcc tct tcc ttt gtc ccc cac aca ctc ttc agg agt cag cca aaa 933  
Asn Ala Ser Ser Phe Val Pro His Thr Leu Phe Arg Ser Gln Pro Lys  
260 265 270

tcc gga tgg tct ttc atg ctg aga att cct gag aag aag aat atg atg 981  
Ser Gly Trp Ser Phe Met Leu Arg Ile Pro Glu Lys Lys Asn Met Met  
275 280 285

tct tcc cgg caa tgg gga cca att ttt ctg aaa gtt ttg cct gga gga 1029  
Ser Ser Arg Gln Trp Gly Pro Ile Phe Leu Lys Val Leu Pro Gly Gly

290                  295                  300                  305  
  
att ttg cag atg tat tat gaa cag gga tta gaa aaa cca ttt aaa gag    1077  
Ile Leu Gln Met Tyr Tyr Glu Gln Gly Leu Glu Lys Pro Phe Lys Glu  
  
310                  315                  320  
  
ata cag ctt gat cca tat tgt agg ctt tct gaa ccc aag gtt gag aac    1125  
Ile Gln Leu Asp Pro Tyr Cys Arg Leu Ser Glu Pro Lys Val Glu Asn  
  
325                  330                  335  
  
ttc agt gta gca gga aaa atc cac act gtg aag att gaa cat gtg tct    1173  
Phe Ser Val Ala Gly Lys Ile His Thr Val Lys Ile Glu His Val Ser  
  
340                  345                  350  
  
tac aca gaa aaa agg aaa tac cat tct aag aca gaa gta gtt cat gaa    1221  
Tyr Thr Glu Lys Arg Lys Tyr His Ser Lys Thr Glu Val Val His Glu  
  
355                  360                  365  
  
cct gac ata gag cag atg ctg aag ttg ggg tcc aca tcg tac cat gac    1269  
Pro Asp Ile Glu Gln Met Leu Lys Leu Gly Ser Thr Ser Tyr His Asp  
370                  375                  380                  385  
  
ttc ctt gac ttt ctg act act gtg gag gag gag ctg atg aag ttg cca    1317  
Phe Leu Asp Phe Leu Thr Thr Val Glu Glu Leu Met Lys Leu Pro  
  
390                  395                  400  
  
gct gtt tca aaa cca aaa aag aac tac gag gag caa gaa att tcc ttg    1365  
Ala Val Ser Lys Pro Lys Lys Asn Tyr Glu Glu Gln Glu Ile Ser Leu  
  
405                  410                  415  
  
gaa att gtg gac aac ttt tgg ggt aaa gtc aca aaa gaa gga aaa ttt    1413  
Glu Ile Val Asp Asn Phe Trp Gly Lys Val Thr Lys Glu Gly Lys Phe  
  
420                  425                  430

gtt gaa agt gct gtg ata act caa att tat tgc ctc tgc ttt gtg aat 1461  
Val Glu Ser Ala Val Ile Thr Gln Ile Tyr Cys Leu Cys Phe Val Asn  
435 440 445

ggg aac ctg gaa tgc ttt tta acc ttg aat gac ctt gag ttg ccg aag 1509  
Gly Asn Leu Glu Cys Phe Leu Thr Leu Asn Asp Leu Glu Leu Pro Lys  
450 455 460 465

cga gat gaa tcc tat tat gag aag gac tca gaa aaa aag ggg att gat 1557  
Arg Asp Glu Ser Tyr Tyr Glu Lys Asp Ser Glu Lys Lys Gly Ile Asp  
470 475 480

att ctt gac tac cat ttt cat aag tgt gtg aat gta caa gaa ttt gag 1605  
Ile Leu Asp Tyr His Phe His Lys Cys Val Asn Val Gln Glu Phe Glu  
485 490 495

caa tca aga atc att aag ttt gta cct ctg gat gcc tgc cgg ttt gag 1653  
Gln Ser Arg Ile Ile Lys Phe Val Pro Leu Asp Ala Cys Arg Phe Glu  
500 505 510

ctg atg cgt ttc aag act ttg tat aat ggg gat aat ctt ccc ttt tcc 1701  
Leu Met Arg Phe Lys Thr Leu Tyr Asn Gly Asp Asn Leu Pro Phe Ser  
515 520 525

ttg aag tct gta gtg gtt gtc cag gga gca tac gtg gaa ctt cag gct 1749  
Leu Lys Ser Val Val Val Val Gln Gly Ala Tyr Val Glu Leu Gln Ala  
530 535 540 545

ttt gtc aac atg gcc tca ttg gcg cag agg tca tcc tat gct ggt tcc 1797  
Phe Val Asn Met Ala Ser Leu Ala Gln Arg Ser Ser Tyr Ala Gly Ser  
550 555 560

tta agg tcc tgt gac aat ata agg ata cac ttt cct gtc cca tcg cag 1845  
Leu Arg Ser Cys Asp Asn Ile Arg Ile His Phe Pro Val Pro Ser Gln

565

570

575

tgg atc aag gcc ctt tgg acc atg aac ctc cag agg cag aag tct ctg 1893  
Trp Ile Lys Ala Leu Trp Thr Met Asn Leu Gln Arg Gln Lys Ser Leu

580

585

590

aaa gct aaa atg aac cgc cga gca tgt ctg ggg agt tta cag gaa ctt 1941  
Lys Ala Lys Met Asn Arg Arg Ala Cys Leu Gly Ser Leu Gln Glu Leu

595

600

605

gaa tct gaa cct gtc att caa gtc act gtg ggg tca gca aaa tat gag 1989  
Glu Ser Glu Pro Val Ile Gln Val Thr Val Gly Ser Ala Lys Tyr Glu  
610 615 620 625

agt gcc tac cag gca gtg gta tgg aag ata gat cgg ctt cca gac aaa 2037  
Ser Ala Tyr Gln Ala Val Val Trp Lys Ile Asp Arg Leu Pro Asp Lys  
630 635 640

aat tca agt cta gat cat ccc cat tgt ctg tca tac aaa tta gag ctt 2085  
Asn Ser Ser Leu Asp His Pro His Cys Leu Ser Tyr Lys Leu Glu Leu  
645 650 655

gga tca gaccaa gaa att ccc tct gat tgg tat cca ttt gct act gtt 2133  
Gly Ser Asp Gln Glu Ile Pro Ser Asp Trp Tyr Pro Phe Ala Thr Val  
660 665 670

cag ttt tcc gtg cct gac acc tgt gcc tca agg aca gag gtc agg tct 2181  
Gln Phe Ser Val Pro Asp Thr Cys Ala Ser Arg Thr Glu Val Arg Ser  
675 680 685

ctg gga gtg gag agt gat gtc cag cca cag aaa cat gtt cag cag cga 2229  
Leu Gly Val Glu Ser Asp Val Gln Pro Gln Lys His Val Gln Gln Arg  
690 695 700 705

gct tgc tac aac atc cag cct aaa ctc tac aga tct gta att gaa gat 2277  
Ala Cys Tyr Asn Ile Gln Pro Lys Leu Tyr Arg Ser Val Ile Glu Asp  
710 715 720

gta att gaa gga gtt cggtt aat cta ttt gct gaa gaa ggt ata gag gaa 2325  
Val Ile Glu Gly Val Arg Asn Leu Phe Ala Glu Glu Gly Ile Glu Glu  
725 730 735

caa gtt tta aaa gac ttg aag cag ctc tgg gaa acc aag gtt ttg cag 2373  
Gln Val Leu Lys Asp Leu Lys Gln Leu Trp Glu Thr Lys Val Leu Gln  
740 745 750

tct aaa gca aca gaa gac ttc ttc aga aat agc atc caa tca cct ctg 2421  
Ser Lys Ala Thr Glu Asp Phe Phe Arg Asn Ser Ile Gln Ser Pro Leu  
755 760 765

ttt act ctt cag ttg ccg cac agc ttg cac caa aca ttg caa tcg tca 2469  
Phe Thr Leu Gln Leu Pro His Ser Leu His Gln Thr Leu Gln Ser Ser  
770 775 780 785

aca gca tca tta gtt att cct gct ggt aga act ctt cca agt ttt acc 2517  
Thr Ala Ser Leu Val Ile Pro Ala Gly Arg Thr Leu Pro Ser Phe Thr  
790 795 800

aca gca gaa ctg ggc act tca aac tcc agt gca aac ttt act ttt cct 2565  
Thr Ala Glu Leu Gly Thr Ser Asn Ser Ala Asn Phe Thr Phe Pro  
805 810 815

ggt tat ccc att cat gta cca gca ggt gtg aca cta cag act gta tct 2613  
Gly Tyr Pro Ile His Val Pro Ala Gly Val Thr Leu Gln Thr Val Ser  
820 825 830

ggt cac ctt tat aaa gtc aat gta cca att atg gtg aca gag act tct 2661  
Gly His Leu Tyr Lys Val Asn Val Pro Ile Met Val Thr Glu Thr Ser

835                  840                  845  
  
gga aga gca ggt att ctt cag cat cca att cag caa gta ttt caa cag    2709  
Gly Arg Ala Gly Ile Leu Gln His Pro Ile Gln Gln Val Phe Gln Gln  
850                  855                  860                  865  
  
ctt ggc cag cct tca gta ata caa act agt gtt cca caa ttg aat cca    2757  
Leu Gly Gln Pro Ser Val Ile Gln Thr Ser Val Pro Gln Leu Asn Pro  
870                  875                  880  
  
tgg tct ctt caa gca act act gaa aaa tca cag aga att gaa acc gtg    2805  
Trp Ser Leu Gln Ala Thr Thr Glu Lys Ser Gln Arg Ile Glu Thr Val  
885                  890                  895  
  
cta cag caa ccc gca att cta cct tct ggg cca gta gat agg aaa cac    2853  
Leu Gln Gln Pro Ala Ile Leu Pro Ser Gly Pro Val Asp Arg Lys His  
900                  905                  910  
  
tta gaa aat gcc acc agt gat ata ctt gta tct cct gga aat gag cat    2901  
Leu Glu Asn Ala Thr Ser Asp Ile Leu Val Ser Pro Gly Asn Glu His  
915                  920                  925  
  
aaa atc gtg cct gaa gct ttg ttg tgt cat cag gaa agt tct cac tat    2949  
Lys Ile Val Pro Glu Ala Leu Leu Cys His Gln Glu Ser Ser His Tyr  
930                  935                  940                  945  
  
atc agt ctt cca ggt gtt gta ttt tct cca cag gtc tct caa aca aat    2997  
Ile Ser Leu Pro Gly Val Val Phe Ser Pro Gln Val Ser Gln Thr Asn  
950                  955                  960  
  
tct gat gtg gag tca gtg ctc agt ggt tca gct agc atg gct caa aat    3045  
Ser Asp Val Glu Ser Val Leu Ser Gly Ser Ala Ser Met Ala Gln Asn  
965                  970                  975

ctg cat gat gag tcc ctc aca agc cct cat ggg gct ctc cac cag 3093  
Leu His Asp Glu Ser Leu Ser Thr Ser Pro His Gly Ala Leu His Gln  
980 985 990

cac gtg act gat att cag ctt cat att ctt aaa aat agg atg tat gga 3141  
His Val Thr Asp Ile Gln Leu His Ile Leu Lys Asn Arg Met Tyr Gly  
995 1000 1005

tgt gat tct gta aag caa cca aga aat ata gag gaa ccc agc aac ata 3189  
Cys Asp Ser Val Lys Gln Pro Arg Asn Ile Glu Glu Pro Ser Asn Ile  
1010 1015 1020 1025

cct gta tca gag aag gat tct aat tct cag gtg gat tta agc att cg 3237  
Pro Val Ser Glu Lys Asp Ser Asn Ser Gln Val Asp Leu Ser Ile Arg  
1030 1035 1040

gtt act gat gat att ggt gaa ata att caa gta gat gga agc ggt 3285  
Val Thr Asp Asp Asp Ile Gly Glu Ile Ile Gln Val Asp Gly Ser Gly  
1045 1050 1055

gat aca tct tcc aat gaa gaa ata gga agt aca aga gat gca gat gag 3333  
Asp Thr Ser Ser Asn Glu Glu Ile Gly Ser Thr Arg Asp Ala Asp Glu  
1060 1065 1070

aat gaa ttt cta ggg aat att gac ggg gga gat ctg aag gta cct gaa 3381  
Asn Glu Phe Leu Gly Asn Ile Asp Gly Asp Leu Lys Val Pro Glu  
1075 1080 1085

gaa gaa gct gac agt att tca aat gag gat tca gcc aca aac agt agt 3429  
Glu Glu Ala Asp Ser Ile Ser Asn Glu Asp Ser Ala Thr Asn Ser Ser  
1090 1095 1100 1105

gat aat gaa gac cct caa gta aac att gta gaa gag gac cct tta aat 3477  
Asp Asn Glu Asp Pro Gln Val Asn Ile Val Glu Asp Pro Leu Asn

1110 1115 1120

tct gga gat gat gtt agt gaa cag gat gtg cca gac ctg ttt gac acg 3525  
Ser Gly Asp Asp Val Ser Glu Gln Asp Val Pro Asp Leu Phe Asp Thr

1125 1130 1135

gat aat gta att gtc tgt cag tat gat aag att cat cga agc aag aac 3573  
Asp Asn Val Ile Val Cys Gln Tyr Asp Lys Ile His Arg Ser Lys Asn

1140 1145 1150

aaa tgg aaa ttc tat ttg aaa gat ggt gtt atg tgt ttt gga ggg aga 3621  
Lys Trp Lys Phe Tyr Leu Lys Asp Gly Val Met Cys Phe Gly Gly Arg

1155 1160 1165

gac tat gta ttt gca aaa gcc att ggt gat gca gag tgg taaaccttgt 3670  
Asp Tyr Val Phe Ala Lys Ala Ile Gly Asp Ala Glu Trp

1170 1175 1180

gagctcagta catctatttt gtgaacatca gttggactat attgcattt gtgaattcat 3730

ttttatTTG aatatAGTCC agcacAGAGC tGTTCAAATT tttAGTTcAC tGTATGGAAT 3790

ttaataaaat tataattcag atgcagatac aatt 3824

<210> 4

<211> 1182

<212> PRT

<213> Homo sapiens

<400> 4

Met Cys Ser Thr Asn Pro Gly Lys Trp Val Thr Phe Asp Asp Asp Pro

1 5 10 15

Ala Val Gln Ser Ser Gln Lys Ser Lys Asn Phe Pro Leu Glu Asn Gln

20

25

30

Gly Val Cys Arg Pro Asn Gly Leu Lys Leu Asn Pro Pro Gly Leu Arg

35

40

45

Glu Phe Pro Ser Gly Ser Ser Ser Thr Ser Ser Thr Pro Leu Ser Ser

50

55

60

Pro Ile Val Asp Phe Tyr Phe Ser Pro Gly Pro Pro Ser Asn Ser Pro

65

70

75

80

Leu Ser Thr Pro Thr Lys Asp Phe Pro Gly Phe Pro Gly Ile Pro Lys

85

90

95

Ala Gly Thr His Val Leu Tyr Pro Ile Pro Glu Ser Ser Ser Asp Ser

100

105

110

Pro Leu Ala Ile Ser Gly Gly Glu Ser Ser Leu Leu Pro Thr Arg Pro

115

120

125

Thr Cys Leu Ser His Ala Leu Leu Pro Ser Asp His Ser Cys Thr His

130

135

140

Pro Thr Pro Lys Val Gly Leu Pro Asp Glu Val Asn Pro Gln Gln Ala

145

150

155

160

Glu Ser Leu Gly Phe Gln Ser Asp Asp Leu Pro Gln Phe Gln Tyr Phe

165

170

175

Arg Glu Asp Cys Ala Phe Ser Ser Pro Phe Arg Lys Asp Glu Gly Ser

180

185

190

Asp Ser His Phe Thr Leu Asp Pro Pro Gly Ser Lys Lys Met Phe Ser

DRAFT COPY - NOT FOR DISTRIBUTION

195

200

205

Ser Arg Asn Lys Glu Met Pro Ile Asp Gln Lys Ser Leu Asn Lys Cys

210

215

220

Ser Leu Asn Tyr Ile Cys Glu Lys Leu Glu His Leu Gln Ser Ala Glu

225

230

235

240

Asn Gln Asp Ser Leu Arg Ser Leu Ser Met His Cys Leu Cys Ala Glu

245

250

255

Glu Asn Ala Ser Ser Phe Val Pro His Thr Leu Phe Arg Ser Gln Pro

260

265

270

Lys Ser Gly Trp Ser Phe Met Leu Arg Ile Pro Glu Lys Lys Asn Met

275

280

285

Met Ser Ser Arg Gln Trp Gly Pro Ile Phe Leu Lys Val Leu Pro Gly

290

295

300

Gly Ile Leu Gln Met Tyr Tyr Glu Gln Gly Leu Glu Lys Pro Phe Lys

305

310

315

320

Glu Ile Gln Leu Asp Pro Tyr Cys Arg Leu Ser Glu Pro Lys Val Glu

325

330

335

Asn Phe Ser Val Ala Gly Lys Ile His Thr Val Lys Ile Glu His Val

340

345

350

Ser Tyr Thr Glu Lys Arg Lys Tyr His Ser Lys Thr Glu Val Val His

355

360

365

Glu Pro Asp Ile Glu Gln Met Leu Lys Leu Gly Ser Thr Ser Tyr His

370

375

380

Asp Phe Leu Asp Phe Leu Thr Thr Val Glu Glu Leu Met Lys Leu  
385                   390                   395                   400

Pro Ala Val Ser Lys Pro Lys Lys Asn Tyr Glu Glu Gln Glu Ile Ser  
405                   410                   415

Leu Glu Ile Val Asp Asn Phe Trp Gly Lys Val Thr Lys Glu Gly Lys  
420                   425                   430

Phe Val Glu Ser Ala Val Ile Thr Gln Ile Tyr Cys Leu Cys Phe Val  
435                   440                   445

Asn Gly Asn Leu Glu Cys Phe Leu Thr Leu Asn Asp Leu Glu Leu Pro  
450                   455                   460

Lys Arg Asp Glu Ser Tyr Tyr Glu Lys Asp Ser Glu Lys Lys Gly Ile  
465                   470                   475                   480

Asp Ile Leu Asp Tyr His Phe His Lys Cys Val Asn Val Gln Glu Phe  
485                   490                   495

Glu Gln Ser Arg Ile Ile Lys Phe Val Pro Leu Asp Ala Cys Arg Phe  
500                   505                   510

Glu Leu Met Arg Phe Lys Thr Leu Tyr Asn Gly Asp Asn Leu Pro Phe  
515                   520                   525

Ser Leu Lys Ser Val Val Val Gln Gly Ala Tyr Val Glu Leu Gln  
530                   535                   540

Ala Phe Val Asn Met Ala Ser Leu Ala Gln Arg Ser Ser Tyr Ala Gly  
545                   550                   555                   560

Ser Leu Arg Ser Cys Asp Asn Ile Arg Ile His Phe Pro Val Pro Ser

565 570 575

Gln Trp Ile Lys Ala Leu Trp Thr Met Asn Leu Gln Arg Gln Lys Ser

580 585 590

Leu Lys Ala Lys Met Asn Arg Arg Ala Cys Leu Gly Ser Leu Gln Glu

595 600 605

Leu Glu Ser Glu Pro Val Ile Gln Val Thr Val Gly Ser Ala Lys Tyr

610 615 620

Glu Ser Ala Tyr Gln Ala Val Val Trp Lys Ile Asp Arg Leu Pro Asp

625 630 635 640

Lys Asn Ser Ser Leu Asp His Pro His Cys Leu Ser Tyr Lys Leu Glu

645 650 655

Leu Gly Ser Asp Gln Glu Ile Pro Ser Asp Trp Tyr Pro Phe Ala Thr

660 665 670

Val Gln Phe Ser Val Pro Asp Thr Cys Ala Ser Arg Thr Glu Val Arg

675 680 685

Ser Leu Gly Val Glu Ser Asp Val Gln Pro Gln Lys His Val Gln Gln

690 695 700

Arg Ala Cys Tyr Asn Ile Gln Pro Lys Leu Tyr Arg Ser Val Ile Glu

705 710 715 720

Asp Val Ile Glu Gly Val Arg Asn Leu Phe Ala Glu Glu Gly Ile Glu

725 730 735

Glu Gln Val Leu Lys Asp Leu Lys Gln Leu Trp Glu Thr Lys Val Leu

740

745

750

Gln Ser Lys Ala Thr Glu Asp Phe Phe Arg Asn Ser Ile Gln Ser Pro

755

760

765

Leu Phe Thr Leu Gln Leu Pro His Ser Leu His Gln Thr Leu Gln Ser

770

775

780

Ser Thr Ala Ser Leu Val Ile Pro Ala Gly Arg Thr Leu Pro Ser Phe

785

790

795

800

Thr Thr Ala Glu Leu Gly Thr Ser Asn Ser Ser Ala Asn Phe Thr Phe

805

810

815

Pro Gly Tyr Pro Ile His Val Pro Ala Gly Val Thr Leu Gln Thr Val

820

825

830

Ser Gly His Leu Tyr Lys Val Asn Val Pro Ile Met Val Thr Glu Thr

835

840

845

Ser Gly Arg Ala Gly Ile Leu Gln His Pro Ile Gln Gln Val Phe Gln

850

855

860

Gln Leu Gly Gln Pro Ser Val Ile Gln Thr Ser Val Pro Gln Leu Asn

865

870

875

880

Pro Trp Ser Leu Gln Ala Thr Thr Glu Lys Ser Gln Arg Ile Glu Thr

885

890

895

Val Leu Gln Gln Pro Ala Ile Leu Pro Ser Gly Pro Val Asp Arg Lys

900

905

910

His Leu Glu Asn Ala Thr Ser Asp Ile Leu Val Ser Pro Gly Asn Glu

915

920

925

His Lys Ile Val Pro Glu Ala Leu Leu Cys His Gln Glu Ser Ser His

930 935 940

Tyr Ile Ser Leu Pro Gly Val Val Phe Ser Pro Gln Val Ser Gln Thr

945 950 955 960

Asn Ser Asp Val Glu Ser Val Leu Ser Gly Ser Ala Ser Met Ala Gln

965 970 975

Asn Leu His Asp Glu Ser Leu Ser Thr Ser Pro His Gly Ala Leu His

980 985 990

Gln His Val Thr Asp Ile Gln Leu His Ile Leu Lys Asn Arg Met Tyr

995 1000 1005

Gly Cys Asp Ser Val Lys Gln Pro Arg Asn Ile Glu Glu Pro Ser Asn

1010 1015 1020

Ile Pro Val Ser Glu Lys Asp Ser Asn Ser Gln Val Asp Leu Ser Ile

1025 1030 1035 1040

Arg Val Thr Asp Asp Ile Gly Glu Ile Ile Gln Val Asp Gly Ser

1045 1050 1055

Gly Asp Thr Ser Ser Asn Glu Glu Ile Gly Ser Thr Arg Asp Ala Asp

1060 1065 1070

Glu Asn Glu Phe Leu Gly Asn Ile Asp Gly Gly Asp Leu Lys Val Pro

1075 1080 1085

Glu Glu Glu Ala Asp Ser Ile Ser Asn Glu Asp Ser Ala Thr Asn Ser

1090 1095 1100

Ser Asp Asn Glu Asp Pro Gln Val Asn Ile Val Glu Glu Asp Pro Leu  
105 1110 1115 1120

Asn Ser Gly Asp Asp Val Ser Glu Gln Asp Val Pro Asp Leu Phe Asp  
1125 1130 1135

Thr Asp Asn Val Ile Val Cys Gln Tyr Asp Lys Ile His Arg Ser Lys  
1140 1145 1150

Asn Lys Trp Lys Phe Tyr Leu Lys Asp Gly Val Met Cys Phe Gly Gly  
1155 1160 1165

Arg Asp Tyr Val Phe Ala Lys Ala Ile Gly Asp Ala Glu Trp  
1170 1175 1180

<210> 5

<211> 261

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)..(30)

<223> Variable 3' end for both ALF and SALF

<400> 5

gca ttc cca aga agg aca tcg ttt aac acc taaactcatt taacaaagga 50

Ala Phe Pro Arg Arg Thr Ser Phe Asn Thr

1 5 10

tccgagaaga acagggacag tgtggagaaga aatcttcttg tcatggcata tttgttcct 110

atatttcttc tggaatcatg ttgcgttggc ttcctgatta aaaacacagt tttattgctc 170

tctgcactgc caaaccaata aatttacaga agagaaagct gtattccact gtacccttg 230

cagcatcaat aaaactgaca gccaaaaaaaa a 261

<210> 6

<211> 10

<212> PRT

<213> Homo sapiens

<400> 6

Ala Phe Pro Arg Arg Thr Ser Phe Asn Thr

1

5

10

<210> 7

<211> 18

<212> DNA

<213> Homo sapiens

<400> 7

agaaaattcccc tctgattg

18

<210> 8

<211> 18

<212> DNA

<213> Homo sapiens

<400> 8

agtaacccga atgcttaa

18

<210> 9  
<211> 18  
<212> DNA  
<213> Homo sapiens

<400> 9  
atgctagctg aaccactg

18

<210> 10  
<211> 27  
<212> DNA  
<213> Homo sapiens

<400> 10  
ccatcctaat acgactcact ataggc

27

<210> 11  
<211> 23  
<212> DNA  
<213> Homo sapiens

<400> 11  
actcaata gggctcgagc ggc

23

<210> 12  
<211> 30  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:oligonucleotide

<400> 12  
ccagaaggta gaattgcggg ttgctgttagc 30

<210> 13  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:oligonucleotide

<400> 13  
ggtgctgtca tggcctgcct caaccgg 28

<210> 14  
<211> 52  
<212> DNA  
<213> Homo sapiens

<400> 14  
actactcata tggcacacca tcaccatcac catgtaccta aactctacag at 52

<210> 15  
<211> 30  
<212> DNA  
<213> Homo sapiens

<400> 15  
agttagtggat ccttaccact ctgcatcacc 30

<210> 16  
<211> 8  
<212> PRT  
<213> Homo sapiens

<400> 16  
Met His His His His His His Val  
1 5

<210> 17  
<211> 6  
<212> PRT  
<213> Homo sapiens

<400> 17  
Met Ala Cys Leu Asn Pro  
1 5

<210> 18  
<211> 23  
<212> DNA  
<213> Homo sapiens

<400> 18  
tactgctcga gcaactttag agt

23

<210> 19  
<211> 29  
<212> DNA  
<213> Homo sapiens

<400> 19

ggagtttcaa gtgcccgagg ctgctgtgg

29

GCTGGAGGTGCTGTCATGGCCTGCCCAACCCGGTGCCTAAACTCTACAGATCTGAATTGAAGATGTAATTGAAGGGAGT  
TCGGAATCTATTTGCTGAGAAGGTTAGAGGAACAAGTTAAAAGACTTGAGCAGCTGGGAAACCAAGGTTTGG  
AGTCTAAAGCAACAGAACAGACTTCTTCAGAAATAGCATCCAATCACCTCTGTTACTCTTCAGTGGCGCACAGCTTGAC  
CAAACATTGCAATGTCACAGCATCATTAGTTATTCCTGCTGTTAGACTCTTCAAGTTTACACAGCAGAACGACTGG  
CACTTCAACATTCCAGTCGAAACTTTACTTTTCTGGTTATCCCCATTCACTGTAACCCAGGTGACACTACAGACTGTT  
CTGGTCACCTTTAAAGTCATGCAATTAGGTCAGAGACAGACTTCTGGAAGAGCAGGTTATCTTCAGCATCCAAATT  
CAGCAAGTATTCAACAGCTTGGCAGCCTTCAGTAATACAAACTAGTGTTCACAAATTGATCCATGGTCTTCAG  
AACTACTGAAAAATCACAGAGAATTGAAACCGTGTACAGCAACCCGAATTCTACCTCTGGGCCAGTAGATAGGAAAC  
ACTTAGAAAATGCCACCAGTGATACTTGTATCTCTGGAAATGAGCATAAAATCGGCCGAAAGCTTGTGTGTCAT  
CAGGAAAGTCTCACTATATCAGTCTCCAGGTGTTGTTATTTCTCACAGGCTCTCAACAAATTCTGATGTGGAGTC  
AGTGTGCTAGTGGTCAGTCAGCATGGCTCAAAATCTGATGAGTCCCTCTCCACAAGCCCTCATGGGCTCTCCACC  
ACCACGTGACTGATATTCTAGCTTCATATTCTAAAAATAGGATGTTAGGATGTTAGCTGTAAGCAACCAAGAAATATA  
GAGGAACCCAGCAACATACCTGTTACAGAGAAGGATTCTAATTCTGGATGTTAGCTGGTTACTGATGATGA  
TATTGGTGAATAATTCAAGTAGATGGAAGCGGTGATACATCTCAATGAGAAATAGGAGTACAAGAGATGCAAGATG  
AGAATGAATTCTAGGAAATTGACGGGGAGATCTGAAGGTACCTGAAAGAAGACGTCAGATTCTCAAATGAGGAT  
TCAGGCCACAAACAGTAGTGATAATTGAAAGACCCCTCAAGTAAACATTGAGAAGAGGACCTTAAATTCTGGAGATGATGT  
TAGTGAAACAGGATGTCAGACGCTGTTGACACGAGATAATTGTAATTGCTGTCAGTATGATAAGATTGATGAAAG  
ACAAATGGAAATTCTATTGAAAGATGGTTATGTTGGAGGGAGAGACTATGTTAGCTGAAAGCCATTGGTGT  
CGAGACAGTGGTAAACCTTGTGAGCTCAGTACATCTATTGTAACATCAGTGGACTATATTGCAATTGTAATTGCAATT  
TTTATTGTAATAGTCCAGCACAGACTGTTCAATTGTTAGTCACTGTTAGTGAATTAAATTATAATTCAAG  
TGCAGATACAAATTACAC

Figure 1.

MACLNPVPKLYRSVIEDVIEGVRNLFAEEGIEEQVLKDLKLQWLTKVLSQSKATEDFFRNSIQSPLFTLQLPHSLHQTLQS  
STASLVIAPAGRTLPSFTAELGTSNSSANFTFPGPYPIHVPAGVTQTVSGHLKVNVPMIVTETSGRAGILQHPIQQVFQ  
QLQCPSPVIQTSPVQLNEWSLQATTEKSQRRIETVLLQQPAILPSPGVDRKHLENATSIDLVSFGNEHKIVPEALLCHQESSH  
YISLPGVVFSPQVSQNTSDVEVSLSGSASMAQNLHDESLSTSFGHALHQHVTDIQLHILKRNVMQGDSVKQPRNIEEPSN  
IPVSEKSDNSQVDSLIRTVPTDDDEGI1QVGDQSGDTSNEEIGSTRDADENEFLGNIDGGDLKVPSEEADSISNEDSATNS  
SDNEDPQVNIVEEDPLNSGDDVSEQDVFDLFDTNVIVCQYDKIHRSKNKWKFYLKDGMCFGGGRDYVFAKIGDAEW

Figure 2.

GGACTTTGGGACTGGACAGACCTGGTACAGTCTAGGTTTACATCTTACTGGTCAGCAACTTAGAGTCACACCTATTT  
 GATTCTTGACAAGACCACAATCTGATCCAAAGATGTGTCACCAAATCCAGGCAATGGGTCACTTGTGATGATGATC  
 CTGCTGTTCAATCTCTAAAAGTCAAGAATTCTCTGGAGAATCAAGGTGTCAGGACCAATGGACTGAAGCTG  
 AACCCCTCTGGCCTCAGGGAATTCCCAGTGGATCTCCTCCACAGCAGCACTCTCTCTCCCTCCCCATTGAGATTT  
 TTATTCAGTCCAGGACCTCCAAGTAACCTCCTCTTCTACACCTACCAAGACTTCCAGGTTTCTGGCATCCCCA  
 AACAGGGACTCATGCTTATCCAGAATCATCTCAGACAGCCCAGCAGCAATATCAGGAGGAGAATCTTCC  
 TTACTGCCTACAGACCAACATGTTATCCATGCCGTTACCAAGTGCACACTCATGTACACATCCAACCTCCAAAGT  
 AGGTCTCCAGTGAAGTAACTCTCACAGGCTGAAGGCAGTCCAAAGTGATGATCTCCCCAGTTCAGTATT  
 TTCGAGAGGACTGTGCTTTCAAGTCATCTCGGAAAGATGAAGGAGTGTGATCTCCATTTCACCCCTGACCCACCAGGA  
 AGCAGAAAAGATGTTCTCATCGAAGAACAGGAGATGCTTATGACCCAAAGGCTAAATAAGTGTCACTCAACTATAT  
 CTGAGAAGCTGAACATCTCAGTCACTGAGAACCAAGACTCACTTAGAAGTTGTCATGCACTGTCTATGCTG  
 AAGAAAATGCCCTCTCCTTGTCCCCACACACTCTCAGGAGTCAGCAAAATCCGGATGCTTCTCATGCTGAGAATT  
 CCTGAGAAGAAGAATATGATGTCCTCCCGCAATGGGGACCAATTCTGAAAGTTGCTGGAGGAATTGAGAT  
 GTATTATGAACAGGATTAGAAAAACATTAAAGAGATACAGCTGATCCATTTGAGCTTCTGAACCCAAAGGTTG  
 AGAACCTCAGTGTAGCAGGAAACACTGTCAGAAGATTGAAGATGTCAGTGTACACAGAAAAAGGAAATACCAATTCT  
 AAGACAGAAGTAGTGTCACTGACATAGAGCAGATGCTGAAGTTGGGTCACACTGTCAGTACTGACTTCTGACTT  
 TCTGACTACTGTGGAGGAGGACCTGATGAAGTCCAGCTGTCAGGCTTCAAAACAAAAGAACACTACAGGAGGAGAAGAAATT  
 CCTTGGAAATTGAGCAACTTTGGGTAAGTCACAAAAGGAAAATTGAGCTGAAAGTGTGATAACTCAAATT  
 TATTGCTCTGTTGTGAATGGGACTTGAATGCTTAACTTGAATGACCTTCTGACTACCATTTCATAAGTGTGTAATGACAAGAAT  
 TTGAGCAATCAAGAATCATTAAGTTGACCTCTGGATGCCCTGCCGTTGAGCTGATGCTTCAAGACTTTGTATAAT  
 GGGGATAATCTCCCTTCTGAAAGTGTGAGTGTGCTGCAAGGAGCATAACGGTAACTCAGGCTTGTCAACAT  
 GGCCATGGGCAAGGGTCACTCTGCTGTTCTGCTGTTCTTAAGGCTCTGTGACAATATAAGGATACACTTCTGCCCCAT  
 CGCAGTGGATCAAGGCCCTTGGACCATGAACTCTGCACTGTCATTCAAGTCACTGTGGGTCAGCAAAATGAACCGCAGGATGT  
 CTGGGAGTTTACAGGAACTTGAACATCTGAACCTGTCACTGTCAGTGTGCTGAGGAGTGTGCTTACCA  
 GCCAGTGGATGAGAAGATAGATGGCTTCAAGCAAAATTCAAGTCTAGATCATCCCCATTGCTGTCAACAAATTAG  
 AGCTTGGATCAGACCAAGAAATTCCCTGATTGATCCATTGCTACTGTCAGTTCCGGTGTGACACCTGTGCC  
 TCAAGGACAGAGGTCAAGGCTCTGGAGTGGAGTGTGTCAGGCCACAGAAACATGTCAGCAGCAGGCTGTACAA  
 CATCCAGCCTAAACTCTACAGATCTGTAATTGAAGATGTAATTGAAGGAGTGTGCAATATTGCTGAAGAAGGTATAG  
 AGGAACAAGTTAAAGACTTGAAGCAGCTGGAAACCAAGTTTGCACTGTCAAAGCAACAGAACAGACTTCTCAGA  
 ATATGATCATTCTGCTGTTACTCTCACTGTCAGGCCACAGCTGCAACAAACATTGCAACAGCATT  
 AGTTATTCTGCTGTTACTCTCAAGTTTCAACAGGACTCTCAAAACAAATTCTGATGTGGAGTGTGCTAGTGTGCAAAACTTACTT  
 TTCCCTGGTTATCCATTCTGTCAGCAGGACTCTCAACAGGACTCTGATGTGTCAGTGTGCACTTAAAGTCAATGTACCA  
 ATTATGGTACAGAGACTCTGGAAAGAGCAGGTATTCTCAGCATCAATTCAAGTCAAGTATTCAACAGCTGGCCAGCC  
 TTCAGTAATACAAACTAGTGTCCACAATTGAATCCATGGTCTCTCAAGCAACTACTGAAAAATCACAGAGAATTGAAA  
 CCGTGTACAGCAACCCGCAATTCTACCTCTGGGCACTAGATAGGAAACACTAGAAAATGCCACCACTGTGATATACTT  
 GTATCTCTGGAATGAGCATAAAATCTGCTGTAAGCTTGTGTCATCAGGAAAGTCTCACTATATCAGTCTTCC  
 AGGTGTTGTTCTCCACAGGTCTCTCAAAACAAATTCTGATGTGGAGTGTGCTAGTGTGTCAGTGTGCT  
 AAAATCTGATGAGTGGCTCTCCACAGGCTCATGGGCTCTCCACAGCAGTGTGACTGATATTGAGCTTCAATATT  
 CTAAAAAATAGGATGATGGATGATCTGTAAGCAACCAAAATAGGAAACCCAGCAACATACCTGTATCAGA  
 GAAGGATTCTAATTCTCAGGTGGATTAAAGCATCGGTTACTGATGATGATATTGGTGAATAATTCAAGTATGGAA  
 GCGGTGATACATCTCCAATGAAGAAATAGGAAAGTACAAGAGATGCAAGATGAGAATGAATTCTAGGGATATTGACGGG  
 GGAGATCTGAAGGTACCTGAAGAAGAGCTGACAGTATTCTAAATGAGGATTGCCACAAACAGTAGTGATAATGAAGA  
 CCCTCAAGTAACATTGAGAAGAGGACCTTAAATTCTGGAGATGATGTTAGTGAACAGGATGTGCCAGACCTGTTG  
 ACACGGATAATGTAATTGTCAGTATGATAAGATTCTGCAAGCAAGAACATGAAATTCTATTGAAAGATGGT  
 GTTATGTGTTGGAGGAGAGACTATGTTGCAAAAGGCTTGGTGTGAGCTGAGTGTGAAACCTTGTGAGGCTCAGTA  
 CATCTATTGTGAACATCATGTTGACTATATTGCAATTGTAATTGAAATTAGTCCAGCACAGAGC  
 GTTCAAATTAGTGTGACTGATGAAATTAAATTCAAGTACATT

**Figure 3.**

MCSTNPWKWTFDDDPAVQSSQSKNFPLENQVCRPNGLKLNPPGLREFPSGSSTSSTPLSSPIVDFYFSPGPPNSP  
 LSTPTKDFPGPGIPKAGTHVLYPIPESSSDSPLAISSGESSLLPRTPTCLSHALLPSDHSCTHPTPKVGLPDEVNPQQA  
 ESLGFQSDDLPQFQYFREDCAFSPFRKDEGSDSHFTLDPPGSKMMSSRNKEMPIDQKSLNKSCLNYICEKLEHLQSAE  
 NQDSLRSLSMHCLCAEENASSFVPHTLFRSQPKSGWSFMLRIPEKKNMMSRQWPFIQLKVLPGGILQMYEQGLEKPK  
 EIQLDPYCRLSEPKVENFSVAGKIHTVKIEHVSYTEKRKYHSKTEVVHEPDIEQMLKLGSTSYHDFLDELTVEEELMKL  
 PAVSKPKKNYEEQEISLEIVDNFWGKVTKEGKFVESAVITQIYCLCFVNGLECFLTLNDLELPKRDESYYEKDSEKKGI  
 DILDYHFHKCVNVQEFEQSRIIKFVPLDACRFELMRFKTLYNGDNLPFSLKSVVVQGAYVELQAFVNMASLAQRSSYAG  
 SLRSCDNIRIHFVPSQWIKALWTMNLQRQKSLKAKMNRRACLGLSLOELESEPVIQVTVGSAKYESAYQAVVWKIDRLPD  
 KNSSLDDHPHCLSYKLELGSDQEIIPSDWYPFATVQFSVPDTCSRTEVRSLVESDVQPQKHVQQRACYNIQPKLYRSVIE  
 DVIEGVNRNLFEEGIEEQVLKDLKQLWETKVLQSKATEDFFRNSIQSPLFTLQLPHSLHQTLQSSTASLVI PAGRTLPSF  
 TTAELGTSNSSANFTFPGYPIHVPAGVTLQTVSGHLYKVNVPIMVETSGRAGILQHPIQOVFQQLGPSPVQIQTSPQLN  
 PWSLQATTEKSQRRIETVLQQPAILPSGPVDRKHLENATSDILVSPGNEHKIVPEALLCHQESSHYIISLPGVVFSPQVSOT  
 NSDVESEVLGSGASMAQNLHDESLSTSHPHALQHVTIDQLHILKNRMYGCDSVKQPRNIEEPSNIPVSEKDSNSQVDLSI  
 RVTDDDIGEIIVQVDGSGDTSNEEIGSTRADENEFLGNIDGGDLKVPPEEADSISNEDSATNNSDNEPDQPVNIVEEDPL  
 NSGDDVSEQDVPDLFDTNVIVCQYDKIHRSKNKWFYKLKGVMFCGGRDYVFAKAIGDAEW

**Figure 4.**

GCATTCCAAGAAGGACATCGTTAACACCTAAACTCATTAAACAAGGATCCGAGAAGAACAGGGACAGTGTGGGAAGA  
AATCTTCTTGATGGCATTTGCTTCTATATTCCTCTGGAAATCATGTTCCCTGGCTTCTGATTAACACAGT  
TTTATTGCTCTGCACTGCCAACCAATAAATTACAGAAGAGAAAGCTGTATTCCACTGTACCCCTGCAGCATCAAT  
AAAACGTGACAGCC

**Figure 5.**

AFPRRTSFNT

**Figure 6.**

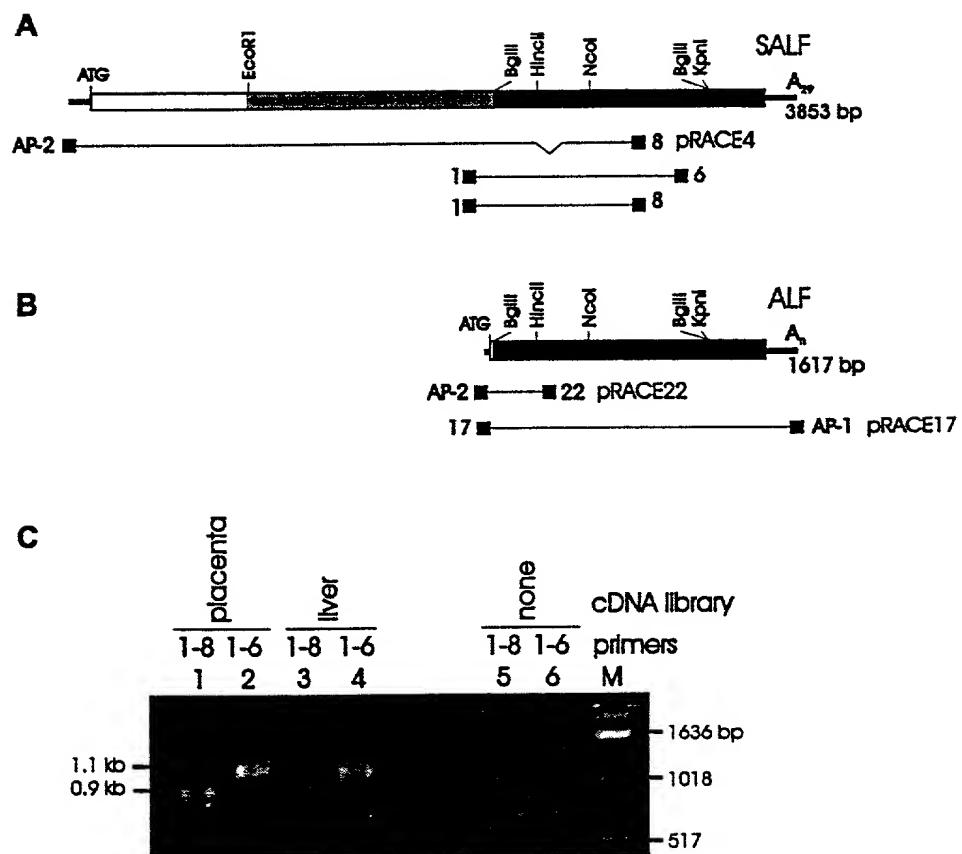
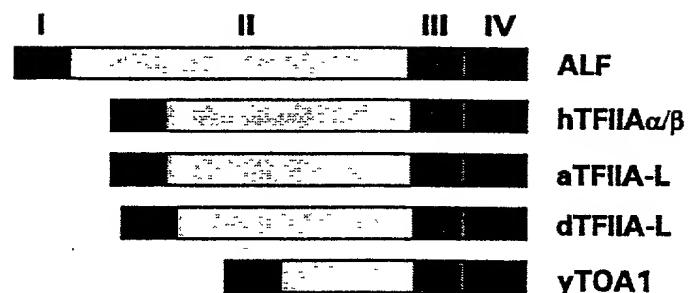


FIG 7 A-C

A



B

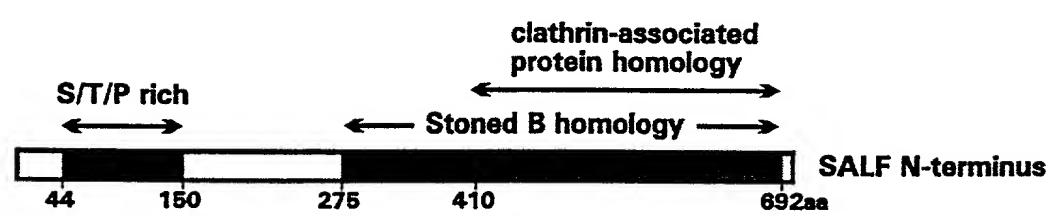


FIG 8A, 8B

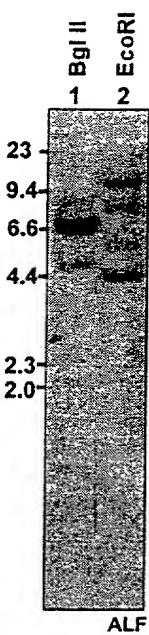


Fig 9

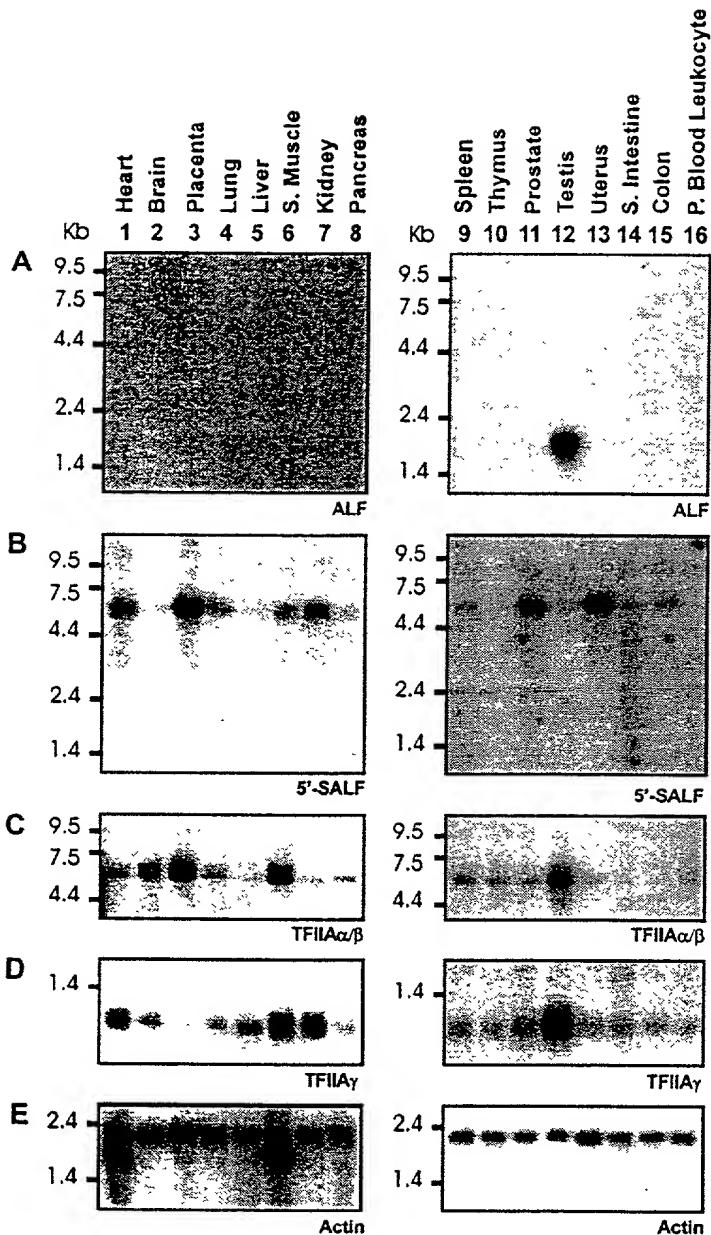


FIG 10 A - E

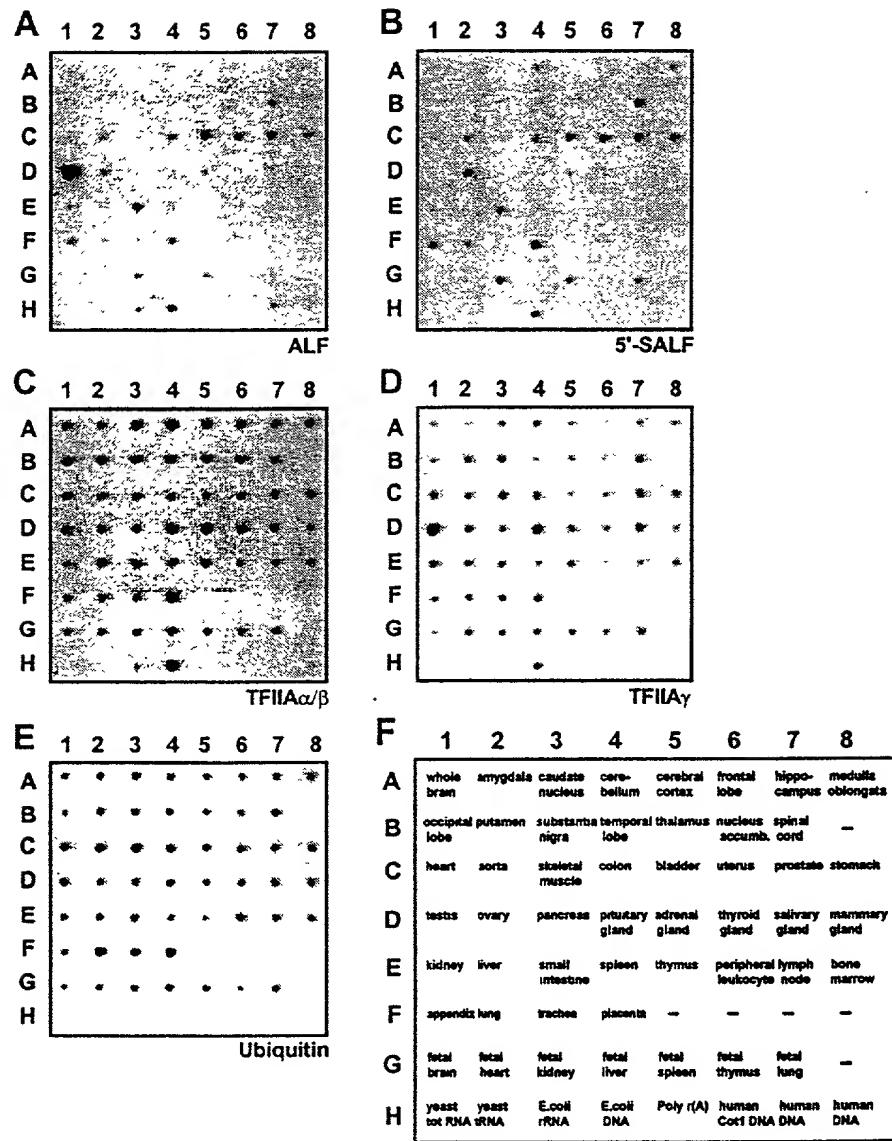


FIG II A - F

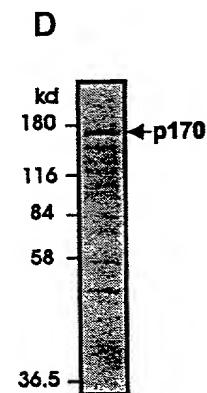
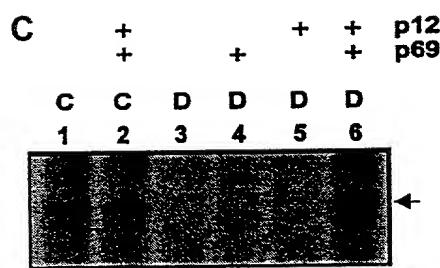
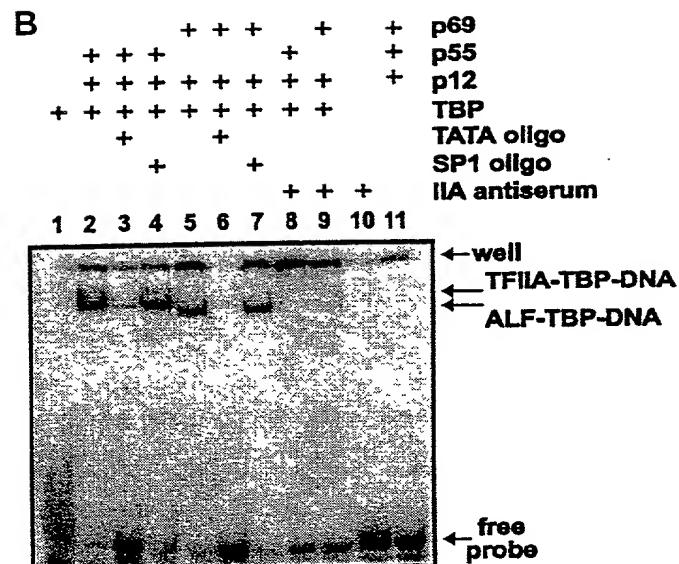
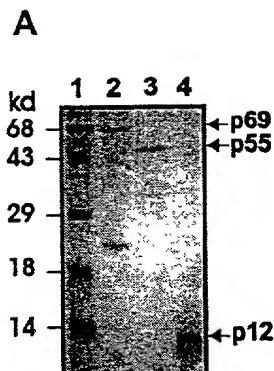


FIG 12 A-D

PATENT  
Attorney Docket No.: 119941-1053  
UTDal No. 99-003

## **DECLARATION**

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe I am the original, first and joint inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled "**Transcription Factors Related to TFIIA**", the Specification of which:

X is attached hereto.  
       was filed on as Application Serial No. .

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims.

I acknowledge the duty to disclose to the Patent and Trademark Office all information known to me to be material to patentability of the subject matter claimed in this application, as "materiality" is defined in Title 37, Code of Federal Regulations, § 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, § 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

### **PRIOR FOREIGN APPLICATION(S)**

Priority	<u>Claimed</u>		
(Number)	(Country)	(Date Filed)	Yes/No

I hereby claim the benefit under 35 U.S.C. §119(e) of any United States provisional application(s) listed below:

None

I hereby claim the benefit under Title 35, United States Code, § 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, § 112, I acknowledge the duty to disclose all information known to me to be material to patentability of the subject matter claimed in this application, as "materiality" is defined in Title 37, Code of Federal Regulations, § 1.56, which become available between the filing date of the prior application and the national or PCT international filing date of this application:

---

(Application Serial No.) (Filing Date) (Status)

---

(Application Serial No.) (Filing Date) (Status)

I hereby direct that all correspondence and telephone calls be addressed to Daniel F. Perez, Gardere & Wynne, L.L.P., 1601 Elm Street, Suite 3000, Dallas, Texas 75201-4761, (214)999-3000.

I hereby declare that all statements made of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Inventor's Full Name: Jeff L. DeJong  
(First) (Initial) (Last)

Inventor's Signature: Jeff DeJong

Date: 8/20/99 Country of Citizenship: USA

Residence Address: 534 Newberry Drive, Richardson, Texas 75080

Post Office Address: \_\_\_\_\_  
(if different from residence address) \_\_\_\_\_

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:

Jeff DeJong

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

§

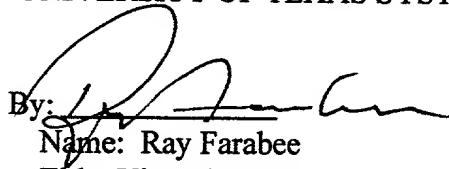
§

Please direct all communications as follows:

Daniel F. Perez, Esq.  
GARDERE & WYNNE, L.L.P.  
3000 Thanksgiving Tower  
1601 Elm Street  
Dallas, Texas 75201-4761

ASSIGNEE:

BOARD OF REGENTS, THE  
UNIVERSITY OF TEXAS SYSTEM

By: 

Name: Ray Farabee  
Title: Vice Chancellor

and General Counsel

Date: May 27, 1999

ASSIGNMENT:

Concurrently filed  
 Previously recorded  
Date: \_\_\_\_\_  
Reel: \_\_\_\_\_  
Frames: \_\_\_\_\_